(11) EP 0 816 499 A2

(12)

## **EUROPEAN PATENT APPLICATION**

(43) Date of publication:07.01.1998 Bulletin 1998/02

(51) Int Cl.<sup>6</sup>: **C12N 15/19**, C07K 14/52, tln 1998/02 C12N 5/10, C07K 1/22

(21) Application number: 97304616.2

(22) Date of filing: 27.06.1997

(84) Designated Contracting States:

AT BE CH DE DK ES FI FR GB GR IE IT LI LU MC

NL PT SE

(30) Priority: 27.06.1996 JP 185305/96

(71) Applicant: KABUSHIKI KAISHA HAYASHIBARA SEIBUTSU KAGAKU KENKYUJO Okayama-shi Okayama (JP)

(72) Inventors:

 Okura, Takanori Kurashiki-shi, Okayama (JP) Torigoe, Kakuji
 Kurashiki-shi, Okayama (JP)

Kurimoto, Masashi
 Okayama-shi, Okayama (JP)

 (74) Representative: Daniels, Jeffrey Nicholas et al Page White & Farrer
 54 Doughty Street London WC1N 2LS (GB)

# (54) Genomic DNA encoding a polypeptide capable of inducing the production of interferon-gamma

(57) Disclosed is a genomic DNA encoding a polypeptide capable of inducing the production of interferon-γ by immunocompetent cells. The genomic DNA efficiently expresses the polypeptide with high biological activities of such as inducing the production of interferon-γ by immunocompetent cells, enhancing killer cells'

cytotoxicity and inducing killer cells' formation, when introduced into mammalian host cells. The high biological activities of the polypeptide facilitate its uses to treat and/or prevent malignant tumors, viral diseases, bacterial infectious diseases and immune diseases without serious side effects when administered to humans.

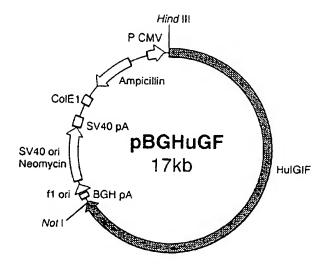


FIG.1

#### Description

The present invention relates to a genomic DNA, more particularly, a genomic DNA encoding a polypeptide capable of inducing the production of interferon-γ (hereinafter abbreviated as "IFN-γ") by immunocompetent cells.

The present inventors successfully isolated a polypeptide capable of inducing the production of IFN- $\gamma$  by immunocompetent cells and cloned a cDNA encoding the polypeptide, which is disclosed in Japanese Patent Kokai No. 27,189/96 and 193,098/96. Because the present polypeptide possesses the properties of enhancing killer cells' cytotoxicity and inducing killer cells' formation as well as inducing IFN- $\gamma$ , a useful biologically active protein, it is expected to be widely used as an agent for viral diseases, microbial diseases, tumors and/or immunopathies, etc.

It is said that a polypeptide generated by a gene expression may be partially cleaved and/or glycosylated by processing with intracellular enzymes in human cells. A polypeptide to be used in therapeutic agents should be preferably processed similarly as in human cells, whereas human cell lines generally have a disadvantage of less producing the present polypeptide, as described in Japanese Patent Application No.269,105/96. Therefore, recombinant DNA techniques should be applied to obtain the present polypeptide in a desired amount. To produce the polypeptide processed similarly as in human cells using recombinant DNA techniques, mammalian cells should be used as the hosts.

In view of foregoing, the first object of the present invention is to provide a DNA which efficiently expresses the polypeptide production when introduced into a mammalian host cell.

The second object of the present invention is to provide a transformant into which the DNA is introduced.

The third object of the present invention is to provide a process for preparing a polypeptide, using the transformant.

#### [Means to Attain the Object]

The present inventors' energetic studies to attain the above objects succeeded in the finding that a genomic DNA encoding the present polypeptide efficiently expresses the polypeptide production when introduced into mammalian host cells. They found that the polypeptide thus obtained possessed significantly higher biological activities than that obtained by expressing a cDNA encoding the polypeptide in *Escherichia*  $\infty li$ .

The first object of the present invention is attained by a genomic DNA encoding a polypeptide with the amino acid sequence of SEQ ID NO:1 (where the symbol "Xaa" means "isoleucine" or "threonine") or its homologous one, which induces interferon-y production by immunocompetent cells.

The second object of the present invention is attained by a transformant formed by introducing the genomic DNA into a mammalian host cell.

The third object of the present invention is attained by a process for preparing a polypeptide, which comprises (a) culturing the transformant in a nutrient medium, and (b) collecting the polypeptide from the resultant culture.

FIG.1 is a restriction map of a recombinant DNA containing a genomic DNA according to the present invention.

Explanation of the symbols are as follows: The symbol "Hin dill" indicates a cleavage site by a restriction enzyme Hin dill, and the symbol "HulGIF" indicates a genomic DNA according to the present invention.

The followings are the preferred embodiments according to the present invention. This invention is made based on the identification of a genomic DNA encoding the polypeptide with the amino acid sequence of SEQ ID NO:1 or its homologous one, and the finding that the genomic DNA efficiently expresses the polypeptide with high biological activities when introduced into mammalian host cells. The genomic DNA of the present invention usually contains two or more exons, at least one of which possesses a part of or the whole of the nucleotide sequence of SEQ ID NO:2. The wording "a part" includes a nucleotide and a sequential nucleotides consisting of two or more nucleotides in SEQ ID NO:2. Examples of the exons are SEQ ID NOs:3 and 4. Human genomic DNA may contain additional exons with SEQ ID NOs:5 to 7. Since the present genomic DNA is derived from a mammalian genomic DNA, it contains introns, as a distinctive feature in mammalian genomic DNAs. The present genomic DNA usually has two or more introns such as SEQ ID NOs:8 to 12.

More particular examples of the present genomic DNA include DNAs with SEQ ID NOs:13 and 14 or complementary sequences thereunto. The DNAs with SEQ ID NOs:13 and 14 are substantially the same. The DNA with SEQ ID NO: 14 contains coding regions for a leader peptide, consisting of the nucleotides 15,607th-15,685th, 17,057th-17,068th and 20,452nd-20,468th, coding regions for the present polypeptide, consisting of the nucleotides 20,469th-20,586th, 21,921st-22,054th and 26,828th-27,046th, and regions as introns, consisting of the nucleotides 15,686th-17,056th, 17,069-20,451st, 20,587th-21,920th and 22,055th-26,827th. The genomic DNA with SEQ ID NO:13 is suitable for expressing the polypeptide in mammalian host cells.

Generally in this field, when artificially expressing a DNA encoding a polypeptide in a host, one or more nucleotides in a DNA may be replaced by different ones, and appropriate promoter(s) and/or enhancer(s) may be linked to the DNA to improve the expressing efficiency or the properties of the expressed polypeptide. The present genomic DNA can be altered similarly as above. Therefore, as far as not substantially changing in the biological activities of the expressed polypeptides, the present genomic DNA should include DNAs encoding functional equivalents of the

20

25

30

5

10

15

35

40

45

55

polypeptide, formed as follows: One or more nucleotides in SEQ ID NOs:3 to 14 are replaced by different ones, the untranslated regions and/or the coding region for a leader peptide in the 5'- and/or 3'-termini of SEQ ID NOs:3, 4, 5, 6, 7, 13 and 14 are deleted, and appropriate oligonucleotides are linked to either or both ends of SEQ ID NO:13.

5

15

20

25

30

35

40

45

50

55

The present genomic DNA includes general DNAs which are derived from a genome containing the nucleotide sequences as above, and it is not restricted to its sources or origins as far as it is once isolated from its original organisms. For example, the present genomic DNA can be obtained by chemically synthesizing based on SEQ ID NOs:2 to 14, or by isolating from a human genomic DNA. The isolation of the present genomic DNA from such a human genomic DNA comprises (a) isolating a genomic DNA from human cells by conventional methods, (b) screening the genomic DNA with probes or primers, which are chemically synthesized oligonucleotides with a part of or the whole of the nucleotide sequence of SEQ ID NO:2, and (c) collecting a DNA to which the probes or primers specifically hybridize. Once the present genomic DNA is obtained, it can be unlimitedly replicated by constructing a recombinant DNA with an autonomously replicable vector by conventional method and then introducing the recombinant DNA into an appropriate host such as a microorganism or an animal cell before culturing the transformant or by applying a PCR method.

The present genomic DNA is very useful in producing the polypeptide by recombinant DNA techniques since it efficiently expresses the polypeptide with high biological activities when introduced into mammalian host cells. The present invention further provides a process for preparing a polypeptide using a specific genomic DNA, comprising the steps of (a) culturing a transformant formed by introducing the present genomic DNA into mammalian host cells, and (b) collecting the polypeptide which induces IFN- $\gamma$  production by immunocompetent cells from the resultant culture.

The following explains the process for preparing the polypeptide according to the present invention. The present genomic DNA is usually introduced into host cells in the form of a recombinant DNA. The recombinant DNA, comprising the present genomic DNA and an autonomously replicable vector, can be relatively easily prepared by conventional recombinant DNA techniques when the genomic DNA is available. The vectors, into which the present genomic DNA can be inserted, include plasmid vectors such as pcD, pcDL-SRa, pKY4, pCDM8, pCEV4 and pME18S. The autonomously replicable vectors usually further contain appropriate nucleotide sequences for the expression of the present recombinant DNA in each host cell, which include sequences for promoters, enhancers, replication origins, transcription termination sites, splicing sequences and/or selective markers. Heat shock protein promoters or IFN-a promoters, as disclosed in Japanese Patent Kokai No.163,368/95 by the same applicant of this invention, enables to artificially regulate the present genomic DNA expression by external stimuli.

To insert the present genomic DNA into vectors, conventional methods used in this field can be arbitrarily used: Genes containing the present genomic DNA and autonomously replicable vectors are cleaved with restriction enzymes and/or ultrasonic, and the resultant DNA fragments and the resultant vector fragments are ligated. To cleave genes and vectors by restriction enzymes, which specifically act on nucleotides, more particularly, *Accl*, *Bam*HI, *Bgl*II, *Bst*XI, *Eco*RI, *Hin*dIII, *Not*I, *Pst*I, *Sac*I, *Sal*I, *Sma*I, *Spe*I, *Xba*I, *Xho*I, etc., facilitate the ligation of the DNA fragments and the vector fragments. To ligate the DNA fragments and the vector fragments, they are, if necessary, first annealed, then treated with a DNA ligase *in vivo* or *in vitro*. The recombinant DNAs thus obtained can be unlimitedly replicated in hosts derived from microorganisms or animals.

Any cells conventionally used as hosts in this field can be used as the host cells: Examples of such are epithelial, interstitial and hemopoietic cells, derived from human, monkey, mouse and hamster, more particularly, 3T3 cells, C127 cells, CHO cells, CV-1 cells, COS cells, HeLa cells, MOP cells and their mutants. Cells which inherently produce the present polypeptide also can be used as the host cells: Example of such are human hemopoietic cells such as lymphoblasts, lymphocytes, monoblasts, monocytes, myeloblasts, myelocytes, granulocytes and macrophages, and human epithelial and interstitial cells derived from solid tumors such as pulmonary carcinoma, large bowel cancer and colon cancer. More particular examples of the latter hemopoietic cells are leukemia cell lines such as HBL-38 cells, HL-60 cells ATCC CCL240, K-562 cells ATCC CCL243, KG-1 cells ATCC CCL246, Mo cells ATCC CRL8066, THP-1 cells ATCC TIB202, U-937 cells ATCC CRL1593.2, described by J. Minowada et al. in "Cancer Research", Vol.10, pp. 1-18 (1988), derived from leukemias or lymphoma including myelogenous leukemias, promyelocytic leukemias, monocytic leukemias, adult T-cell leukemias and hairy cell leukemias, and their mutants. The present polypeptide-processibility of these leukemia cell lines and their mutants is so distinguished that they can easily yield the polypeptide with higher biological activities when used as hosts.

To introduce the present DNA into the hosts, conventional methods such as DEAE-dextran method, calcium phosphate transfection method, electroporation method, lipofection method, microinjection method, and viral infection method as using retrovirus, adenovirus, herpesvirus and vaccinia virus, can be used. The polypeptide-producing clones in the transformants can be selected by applying the colony hybridization method or by observing the polypeptide production after culturing the transformants in culture media. For example, the recombinant DNA techniques using mammalian cells as hosts are detailed in "Jikken-Igaku-Bessatsu Saibo-Kogaku Handbook (The handbook for the cell engineering)" (1992), edited by Toshio KUROKI, Masaru TANIGUCHI and Mitsuo OSHIMURA, published by YODOSHA. CO., LTD., Tokyo, Japan, and "Jikken-Igaku-Bessatsu Biomanual Series 3 Idenshi Cloning Jikken-Ho (The experimen-

tal methods for the gene cloning)" (1993), edited by Takahi YOKOTA and Ken-ichi ARAI, published by YODOSHA CO., LTD., Tokyo, Japan.

5

15

20

25

35

45

50

The transformants thus obtained secrete the present polypeptide intracellularly and/or extracellularly when cultured in culture media. As the culture media, conventional ones used for mammalian cells can be used. The culture media generally comprise (a) buffers as a base, (b) inorganic ions such as sodium ion, potassium ion, calcium ion, phosphoric ion and chloric ion, (c) micronutrients, carbon sources, nitrogen sources, amino acids and vitamins, which are added depending on the metabolic ability of the cells, and (d) sera, hormones, cell growth factors and cell adhesion factors, which are added if necessary. Examples of individual media include 199 medium, DMEM medium, Ham's F12 medium, IMDM medium, MCDB 104 medium, MCDB 153 medium, MEM medium, RD medium, RITC 80-7 medium, RPMI-1630 medium, RPMI-1640 medium and WAJC 404 medium. The cultures containing the present polypeptide are obtainable by inoculating the transformants into the culture media to give a cell density of 1 x 10<sup>4</sup> - 1 x 10<sup>7</sup> cells/ml, more preferably, 1 x 10<sup>5</sup> - 1 x 10<sup>6</sup> cells/ml, and then subjecting to suspension- or monolayer-cultures at about 37°C for 1-7 days, more preferably, 2-4 days, while appropriately replacing the culture media with a fresh preparation of the culture media. The cultures thus obtained usually contain the present polypeptide in a concentration of about 1-100 μg/ml, which may vary depending on the types of the transformants or the culture conditions used.

While the cultures thus obtained can be used intact as an IFN-y inducer, they are usually subjected to a step for separating the present polypeptide from the cells or the cell debris using filtration, centrifugation, etc. before use, which may follow a step for disrupting the cells with supersonication, cell-lytic enzymes and/or detergents if desired, and to a step for purifying the polypeptide. The cultures from which the cells or cell debris are removed are usually subjected to conventional methods used in this field for purifying biologically active polypeptides, such as salting-out, dialysis, filtration, concentration, separatory sedimentation, ion-exchange chromatography, gel filtration chromatography, adsorption chromatography, chromatofocusing, hydrophobic chromatography, reversed phase chromatography, affinity chromatography, gel electrophoresis and/or isoelectric focusing. The resultant purified polypeptide can be concentrated and/or lyophilized into liquids or solids depending on final uses. The monoclonal antibodies disclosed in Japanese Patent Kokai No.231,598/96 by the same applicant of this invention are extremely useful to purify the present polypeptide. Immunoaffinity chromatography using monoclonal antibodies yields the present polypeptide in a relatively high purity at the lowest costs and labors.

The polypeptide obtainable by the process according to the present invention exerts strong effects in the treatment and/or the prevention for IFN- $\gamma$ - and/or killer cells-usceptive diseases since it possesses the properties of enhancing killer cells' cytotoxicity and inducing killer cells' formation as well as inducing IFN- $\gamma$ , a useful biologically active protein, as described above. The polypeptide according to the present invention has a high activity of inducing IFN- $\gamma$ , and this enables a desired amount of IFN- $\gamma$  production with only a small amount. The polypeptide is so low toxic that it scarcely causes serious side effects even when administered in a relatively-high dose. Therefore, the polypeptide has an advantage that it can readily induce IFN- $\gamma$  in a desired amount without strictly controlling the dosage. The uses as agents for susceptive diseases are detailed in Japanese Patent Application No.28,722/96 by the same applicant of this invention.

The present genomic DNA is also useful for so-called "gene therapy". According to conventional gene therapy, the present DNA can be introduced into patients with IFN-γ- and/or killer cell-susceptive diseases by directly injecting after the DNA is inserted into vectors derived from viruses such as retrovirus, adenovirus and adeno-associated virus or is incorporated into cationic- or membrane fusible-liposomes, or by self-transplanting lymphocytes which are collected from patients before the DNA is introduced. In adoptive immunotherapy with gene therapy, the present DNA is introduced into effector cells similarly as in conventional gene therapy. This can enhance the cytotoxicity of the effector cells to tumor cells, resulting in improvement of the adoptive immunotherapy. In tumor vaccine therapy with gene therapy, tumor cells from patients, into which the present genomic DNA is introduced similarly as in conventional gene therapy, are self-transplanted after proliferated *ex vivo* up to give a desired cell number. The transplanted tumor cells act as vaccines in the patients to exert a strong antitumor immunity specifically to antigens. Thus, the present genomic DNA exhibits considerable effects in gene therapy for diseases including viral diseases, microbial diseases, malignant tumors and immunopathies. The general procedures for gene therapy are detailed in "Jikken-Igaku-Bessatsu Biomanual UP Series Idenshichiryo-no-Kisogijutsu (Basic techniques for the gene therapy)" (1996), edited by Takashi ODA-JIMA, Izumi SAITO and Keiya OZAWA, published by YODOSHA CO., LTD., Tokyo, Japan.

The following examples explain the present invention, and the techniques used therein are conventional ones used in this field: For example, the techniques are described in "Jikken-Igaku-Bessatsu Saibo-Kogaku Handbook (The handbook for the cell engineering)", (1992), edited by Toshio KUROKI, Masaru TANIGUCHI and Mitsuo OSHIMURA, published by YODOSHA CO., LTD., Tokyo, Japan, and "Jikken-Igaku-Bessatsu Biomanual Series 3 Idenshi Clonong Jikken-Ho (The experimental methods for the gene cloning)" (1993), edited by Takahi YOKOTA and Ken-ichi ARAI, published by YODOSHA CO., LTD., Tokyo, Japan.

#### Example 1

## Cloning genomic DNA and determination of nucleotide sequence

#### 5 Example 1-1

15

20

25

30

35

40

45

#### Determination of partial nucleotide sequence

Five ng of "PromoterFinder™ DNA Pvull LIBRARY", a human placental genomic DNA library commercialized by CLONTECH Laboratories, Inc., California, USA, 5 μl of 10 x Tth PCR reaction solution, 2.2 μl of 25 mM magnesium acetate, 4 μl of 2.5 mM dNTP-mixed solution, one μl of the mixed solution of 2 unit/μl rTth DNA polymerase XL and 2.2 μg/μl Tth Start Antibody in a ratio of 4:1 by volume, 10 pmol of an oligonucleotide with the nucleotide sequence of 5'-CCATCCTAATACGACTCACTATAGGGC-3' as an adaptor primer, and 10 pmol of an oligonucleotide with the nucleotide sequence of 5'-TTCCTCTTCCCGAAGCTGTAGACTGC-3' as an anti-sense primer, which was chemically synthesized based on the sequence of the nucleotides 88th-115th in SEQ ID NO:2, were mixed and volumed up to 50 μl with sterilized distilled water. After incubating at 94°C for one min, the mixture was subjected to 7 cycles of incubations at 94°C for 25 sec and at 72°C for 4 min, followed by 32 cycles of incubations at 94°C for 25 sec at 67°C for 4 min to perform PCR.

The reaction mixture was diluted by 100 folds with sterilized distilled water. One μl of the dilution, 5 μl of 10 x Tth PCR reaction solution, 2.2 μl of 25 mM magnesium acetate, 4 μl of 2.5 mM dNTP-mixed solution, one μl of the mixed solution of 2 unit/μl rTth DNA polymerase XL and 2.2 μg/μl Tth Start Antibody in a ratio of 4:1 by volume, 10 pmol of an oligonucleotide with the nucleotide sequence of 5'-CTATAGGCACGCGTGGT-3' as a nested primer, and 10 pmol of an oligonucleotide with the nucleotide sequence of 5'-TTCCTCTTCCCGAAGCTGTAGACTGC-3' as an antisense primer, which was chemically synthesized similarly as above, were mixed and volumed up to 50 μl with sterilized distilled water. After incubating at 94°C for one min, the mixture was subjected to 5 cycles of incubations at 94°C for 25 sec and at 72°C for 4 min, followed by 22 cycles of incubations at 94°C for 25 sec and at 67°C for 4 min to perform PCR for amplifying a DNA fragment of the present genomic DNA. The genomic DNA library and reagents for PCR used above were mainly from "PromoterFinder™ DNA WALKING KITS", commercialized by CLONTECH Laboratories, Inc., California, USA

An adequate amount of the PCR product thus obtained was mixed with 50 ng of "pT7 Blue(R)", a plasmid vector commercialized by Novagen, Inc., WI, USA, and an adequate amount of T4 DNA ligase, and 100 mM ATP was added to give a final concentration of one mM, followed by incubating at 16°C for 18 hr to insert the DNA fragment into the plasmid vector. The obtained recombinant DNA was introduced into an *Escherichia coli* JM109 strain by the competent cell method to form a transformant, which was then inoculated into L-broth medium (pH 7.2) containing 50 µg/ml ampicillin and cultured at 37°C for 18 hr. The cells were isolated from the resulting culture, and then subjected to the conventional alkali-SDS method to collect a recombinant DNA. The dideoxy method analysis confirmed that the recombinant DNA contained the DNA fragment with a sequence of the nucleotides 5,150th-6,709th in SEQ ID NO:14.

#### Example 1-2

#### Determination of partial nucleotide sequence

PCR was performed in the same conditions as the first PCR in Example 1-1, but an oligonucleotide with the nucleotide sequence of 5'-GTAAGTTTTCACCTTCCAACTGTAGAGTCC-3', which was chemically synthesized based on the nucleotide sequence of the DNA fragment in Example 1-1, was used as an anti-sense primer.

The reaction mixture was diluted by 100 folds with sterilized distilled water. One µl of the dilution was placed into a reaction tube, and PCR was performed in the same conditions as used in the second PCR in Example 1-1 to amplify another DNA fragment of the present genomic DNA, but an oligonucleotide with the nucleotide sequence of 5'-GGGAT-CAGATCAGAAGCAGCACAC-3', which was chemically synthesized based on the nucleotide sequence of the DNA fragment in Example 1-1, was used as an anti-sense primer.

The DNA fragment was inserted into the plasmid vector similarly as in Example 1-1 to obtain a recombinant DNA. The recombinant DNA was replicated in *Escherichia coli* before being collected. The analysis of the collected recombinant DNA confirmed that it contained the DNA fragment with a sequence of the nucleotides lst-5,228th in SEQ ID NO:14.

55

#### Example 1-3

5

15

30

35

#### Determination of partial nucleotide sequence

0.5 μg of a human placental genomic DNA, commercialized by CLONTECH Laboratories, Inc., California, USA, 5 μl of 10 x PCR reaction solution, 8 μl of 2.5 mM dNTP-mixed solution, one μl of the mixed solution of 5 unit/μl "TAKARA LA Taq POLYMERASE" and 1.1 μg/μl "TaqStart ANTIBODY" in a ratio of 1:1 by volume, both of them are commercialized by Takara Syuzo Co., Tokyo, Japan, 10 pmol of an oligonucleotide with the nucleotide sequence of 5'-CCTGGCT-GCCAACTCTGGCTAAAGCGG-3' as a sense primer, chemically synthesized based on a sequence of the nucleotides 46th-75th in SEQ ID NO:2, and 10 pmol of an oligonucleotide with the nucleotide sequence of 5'-GTATTGT-CAATAAATTTCATTGCCACAAAGTTG-3' as an anti-sense primer, chemically synthesized based on a sequence of the nucleotides 210th-242nd in SEQ ID NO:2, were mixed and volumed up to 50 μl with sterilized distilled water. After incubating at 94°C for one min, the mixture was subjected to 5 cycles of incubations at 98°C for 20 sec and at 68°C for 10 min, followed by 25 cycles of incubations at 98°C for 20 sec and 68°C for 10 min, with adding 5 sec in times to every cycle, and finally incubated at 72°C for 10 min to amplify further DNA fragment of the present genomic DNA. The reagents for PCR used above were mainly from "TAKARA LA PCR KIT VERSION 2", commercialized by Takara Syuzo Co., Tokyo, Japan.

The DNA fragment was inserted into the plasmid vector similarly as in Example 1-1 to obtain a recombinant DNA. The recombinant DNA was replicated in *Escherichia coli* before being collected. The analysis of the collected recombinant DNA confirmed that it contained the DNA fragment with a sequence of the nucleotides 6,640th-15,671st in SEQ ID NO:14.

#### Experiment 1-4

#### 25 Determination of partial nucleotide sequence

PCR was performed in the same conditions as the PCR in Example 1-3 to amplify further another DNA fragment of the present genomic DNA; but an oligonucleotide with the nucleotide sequence of 5'-AAGATGGCTGCAACCAGTAGAAGACAATTGC-3', chemically synthesized based on a sequence of the nucleotide 175th-207th in SEQ ID NO: 2, was used as a sense primer, an oligonucleotide with the nucleotide sequence of 5'-TCCTTGGTCAATGAAGAGACCTTGGTC-3', chemically synthesized based on a sequence of nucleotides 334th-360th in the SEQ ID NO:2, was used as an anti-sense primer, and after incubating at 98°C for 20 sec, the reaction mixture was subjected to 30 cycles of incubations at 98°C for 20 sec and at 68°C for 3 min, followed by incubating at 72°C for 10 min.

The DNA fragment was inserted into the plasmid vector similarly as in Example 1-1 to obtain a recombinant DNA. The recombinant DNA was replicated in *Escherichia coli* before being collected. The analysis of the collected recombinant DNA confirmed that it contained the DNA fragment with a sequence of the nucleotides 15,604th-20,543rd in SEQ ID NO:14.

#### Example 1-5

### 40

## Determination of partial nucleotide sequence

PCR was performed in the same conditions as the PCR in Example 1-4 to amplify further another DNA fragment of the present genomic DNA, but an oligonucleotide with the nucleotide sequence of 5'-CCTGGAATCAGATTACTTT-GGCAAGCTTGAATC-3', chemically synthesized based on the sequence of the nucleotide 273rd-305th in SEQ ID NO: 2, was used as a sense primer, and an oligonucleotide with the nucleotide sequence of 5'-GGAAATAATTTTGTTCT-CACAGGAGAGAGTTG-3', chemically synthesized based on the sequence of nucleotides 500th-531st in the SEQ ID NO:2, was used as an anti-sense primer.

The DNA fragment was inserted into the plasmid vector similarly as in Example 1-1 to obtain a recombinant DNA. The recombinant DNA was replicated in *Escherichia coli* before being collected. The analysis of the collected recombinant DNA confirmed that it contained the DNA fragment with a sequence of the nucleotides 20,456th-22,048th in SEQ ID NO:14.

#### Example 1-6

#### 55

50

#### Determination of partial nucleotide sequence

PCR was performed in the same conditions as the PCR in Example 1-4 to amplify further another DNA fragment

of the present genomic DNA, but an oligonucleotide with the nucleotide sequence of 5'-GCCAGCCTAGAGGTATGGCT-GTAACTATCTC-3', chemically synthesized based on the sequence of the nucleotide 449th-479th in SEQ ID NO:2, was used as a sense primer, and an oligonucleotide with the nucleotide sequence of 5'-GGCATGAAATTTTAAT-AGCTAGTCTTCGTTTTG-3', chemically synthesized based on the sequence of nucleotides 745th-777th in the SEQ ID NO:2, was used as an anti-sense primer.

The DNA fragment was inserted into the plasmid vector similarly as in Example 1-1 to obtain a recombinant DNA. The recombinant DNA was replicated in *Escherichia coli* before being collected. The analysis of the collected recombinant DNA confirmed that it contained the DNA fragment with a sequence of the nucleotides 21,996th-27,067th in SEQ ID NO:14.

Example 1-7

5

10

15

20

25

30

40

50

#### Determination of partial nucleotide sequence

PCR was performed in the same conditions as the first PCR in Example 1-2 to amplify further another DNA fragment in the present genomic DNA, but an oligonucleotide with the nucleotide sequence of 5'-GTGACATCATATTCTTTCA-GAGAAGTGTCC-3', chemically synthesized based on the sequence of the nucleotide 575th-604th in SEQ ID NO:2, was used as a sense primer.

The reaction mixture was diluted by 100 folds with sterilized distilled water. One µI of the dilution was placed into a reaction tube, and PCR was performed in the same conditions as the second PCR in Example 1-2 to amplify further another DNA fragment of the present genomic DNA, but an oligonucleotide with the sequence of 5'-GCAATTTGAATCT-TCATCATACGAAGGATAC-3', chemically synthesized based on a sequence of the nucleotides 624th-654th in SEQ ID NO:2, was used as a sense primer.

The DNA fragment was inserted into the plasmid vector similarly as in Example 1-1 to obtain a recombinant DNA. The recombinant DNA was replicated in *Escherichia coli* before being collected. The analysis of the collected recombinant DNA confirmed that it contained the DNA fragment with a sequence of the nucleotides 26,914th-28,994th in SEQ ID NO:14.

#### Example 1-8

#### Determination of complete nucleotide sequence

Comparing the nucleotide sequence of SEQ ID NO:2, which was proved to encode the present polypeptide, as disclosed in Japanese Patent Kokai No.193,098/96 by the same applicant of this invention, with the partial nucleotide sequences identified in Examples 1-1 to 1-7, it was proved that the present genomic DNA contained the nucleotide sequence of SEQ ID NO:14. SEQ ID NO:14, consisting of 28,994 base pairs (bp), was extremely longer than the SEQ ID NO:2, consisting of only 471 bp. This suggested that SEQ ID NO:14 contained introns, a characteristic of eukalyotic cells.

It was examined where partial nucleotide sequences of SEQ ID NO:2, i.e., exons, and the donor and acceptor sites in introns, respectively consisting of the nucleotides of GT and AG, located in SEQ ID NO:14. Consequently, it was proved that SEQ ID NO:14 contained at least 5 introns, which located in the order of SEQ ID NOs:10, 11, 12, 8 and 9 in the direction from the 5'- to the 3'-termini. Therefore, the sequences between the neighboring introns must be exons, which were thought to be located in the order of SEQ ID NOs:5, 6, 3, 4 and 7 in the direction from the 5'- to the 3'-termini. It was also proved that SEQ ID NO:7 contained the 3'-untranslated region other than the exons. The features of the sequence elucidated as this are arranged in SEQ ID NO:14.

As disclosed in Japanese Patent Application No. 269,105/96 by the same applicant of this invention, the present polypeptide is produced as a polypeptide with N-terminal amino acid of tyrosine other than methionine in human cells, which is observed in SEQ ID NO:1. This suggests that the present genomic DNA contains a leader peptide region in the upstream of the 5'-terminus of the present polypeptide-encoding region. A sequence consisting of 36 amino acids encoded by the upstream of the nucleotides 20,469th-20,471st. which is the nucleotides of TAC, are described as a leader peptide in SEQ ID NO:14.

#### Example 2

## 55 Preparation of recombinant DNA pBGHuGF for expression

0.06 ng of the DNA fragment in Example 1-4 in a concentration of 3 ng/50 μl, 0.02 ng of the DNA fragment, obtained by the methods in Example 1-5, 5 μl of 10 x LA PCR reaction solution, 8 μl of 2.5 mM dNTP-mixed solution, one μl of

the mixed solution of 5 unit/µl TAKARA LA Taq polymerase and 1.1 µg/µl TaqStart Antibody in a ratio of 1:1 by volume, 10 pmol of an oligonucleotide with the sequence of 5'-TCCGAAGCTTAAGATGGCTGCAACCAGTA-3' as a sense primer, chemically synthesized based on the nucleotide sequence of the DNA fragment in Example 1-4, and 10 pmol of an oligonucleotide with the nucleotide sequence of 5'-GGAAATAATTTTGTTCTCACAGGAGAGAGTTG-3' as an anti-sense primer, chemically synthesized based on the nucleotide sequence of the DNA fragment in Example 1-5, were mixed and volumed up to 50 µl with sterilized distilled water. After incubating at 94°C for one min, the mixture was subjected to 5 cycles of incubations at 98°C for 20 sec and at 72°C for 7 min, followed by 25 cycles of incubations at 98°C for 20 sec and 68°C for 7 min to perform PCR. The reaction mixture was cleaved by restriction enzymes *Hin*dIII and *Sph*I to obtain a DNA fragment of about 5,900 bp, with cleavage sites by HindIII and *Sph*I in its both termini.

PCR was performed in the same condition as above, but 0.02 ng of the DNA fragment in Example 1-5, 0.06 ng of the DNA fragment obtained in Example 1-6, an oligonucleotide with the nucleotide sequence of 5'-ATGTAGCG-GCCGCGCATGAAATTTTAATAGCTAGTC-3' as an anti-sense primer, chemically synthesized based on the nucleotide sequence of the DNA fragment in Example 1-6, and an oligonucleotide with the sequence of 5'-CCTGGAATCA-GATTACTTTGGCAAGCTTGAATC-3' as a sense primer, chemically synthesized based on the DNA fragment in Example 1-6, were used. The reaction mixture was cleaved by restriction enzymes *Not*I and *Sph*I to obtain a DNA fragment of about 5,600 bp, with cleavage sites by *Not*I and *Sph*I in its both termini.

A plasmid vector "pRc/CMV", containing a *cytomegalovirus* promoter, commercialized by Invitrogen Corporation, San Diego, USA, was cleaved by restriction enzymes *Hind*III and *Not*I to obtain a vector fragment of about 5,500 bp. The vector fragment was mixed with the above two DNA fragments of about 5,900 bp and 5,600 bp, and reacted with T4 DNA ligase to insert the two DNA fragments into the plasmid vector. An *Escherichia coli* JM109 strain was transformed with the obtained recombinant DNA, and the transformant with the plasmid vector was selected by the colony hybridization method. The selected recombinant DNA was named as "pBGHuGF". As shown in FIG.1, the present genomic DNA, with the nucleotide sequence of SEQ ID NO:13, was ligated in the downstream of the cleavage site by the restriction enzyme HindIII in the recombinant DNA.

#### Example 3

10

15

20

25

35

40

45

#### Preparation of transformant using CHO cell as host

CHO-K1 cells ATCC CCL61 were inoculated into Ham's F12 medium (pH 7.2) containing 10 v/v % bovine fetal serum and proliferated by conventional manner. The proliferated cells were collected and washed with phosphate-buffered saline (hereinafter abbreviated as "PBS") followed by suspending in PBS to give a cell density of 1 x 10<sup>7</sup> cells/ml.

10 μg of the recombinant DNA pBGHuGF in Example 2 and 0.8 ml of the above cell suspension were placed in a cuvette and ice-chilled for 10 min. The cuvette was installed in "GENE PULSER", an electroporation device commercialized by Bio-Rad Laboratories Inc., Brussels, Belgium, and then pulsed once with an electric discharge. After pulsing, the cuvette was immediately took out and ice-chilled for 10 min. The cell suspension from the cuvette was inoculated into Ham's F12 medium (pH 7.2) containing 10 v/v % bovine fetal serum and cultured under an ambient condition of 5 v/v % CO<sub>2</sub> at 37°C for 3 days. To the culture medium was added G-418 to give a final concentration of 400 μg/ml, and the culturing was continued further 3 weeks under the same conditions. From abut 100 colonies formed, 48 colonies were selected, and a portion of each was inoculated into a well of culturing plates with Ham's F12 medium (pH7.2) containing 400 μg/ml G-418 and 10 v/v % bovine fetal serum and cultured similarly as above. Thereafter, to each well of the culturing plates was added 10 mM Tris-HCl buffer (pH 8.5) containing 5.1 mM magnesium chloride, 0.5 w/v % sodium deoxycholate, 1 w/v % NONIDET P-40, 10 μg/ml aprotinin and 0.1 w/v % SDS to lyse the cells.

50 µl aliquot of the cell lysates was mixed with one ml of glycerol and incubated at 37°C for one hour, before the polypeptides in the cell lysates were separated by the SDS-polyacrylamide gel electrophoresis. The separated polypeptides were transferred to a nitrocellulose membrane in usual manner, and the membrane was soaked in the culture supernatant of the hybridoma H-1, disclosed in Japanese Patent Kokai No.231,598/96 by the same applicant of this invention, followed by washing with 50 mM Tris-HCl buffer containing 0.05 v/v % TWEEN 20 to remove an excessive mount of the monoclonal antibody. Thereafter, the nitrocellulose membrane was soaked in PBS containing rabbit-derived anti-mouse immunoglobulin antibody for one hr, which was labeled with horseradish peroxidase, followed by washing 50 mM Tris-HCl buffer (pH 7.5) containing 0.05 v/v % TWEEN 20 and soaking in 50 mM Tris-HCl buffer (pH 7.5) containing 0.005 v/v % hydrogen peroxide and 0.3 mg/ml diaminobenzidine to develop colorations. The clone, which highly produced the polypeptide, was selected based on the color development and named "BGHuGF".

55

#### Example 4

5

15

20

25

30

#### Production of polypeptide by transformant and its physicochemical properties

The transformant BGHuGF in Experiment 3 was inoculated into Ham's F12 medium (pH 7.2) containing 400 µg/ml G-418 and 10 v/v % bovine fetal serum, and cultured under an ambient condition of 5 v/v % CO<sub>2</sub> at 37°C for one week. The proliferated cells were collected, washed with PBS, and then washing with 10-fold volumes of ice-chilled 20 mM Hepes buffer (pH 7.4), containing 10 mM potassium chloride and 0.1 mM ethylendiaminetetraacetate bisodium salt, according to the method described in "Proceedings of The National Academy of The Sciences of The USA", vol. 86, pp.5,227-5,231 (1989), by M. J. Kostura et al. The cells thus obtained were allowed to stand in 3-fold volumes of a fresh preparation of the same buffer under an ice-chilling condition for 20 min and freezed at -80°C, succeeded by thawing to disrupt the cells. The resulting cells were centrifuged to collect the supernatant.

In parallel, THP-1 cells ATCC TIB 202, derived from a human acute monocytic leukemia, was similarly cultured and disrupted. Supernatant, obtained by centrifuging the resulting cells, was mixed with the supernatant obtained from the transformant BGHuGF and incubated at 37°C for 3 hr to react. The reaction mixture was applied to a column with "DEAE-SEPHAROSE", a gel for ion-exchange chromatography, commercialized by Pharmacia LKB Biotechnology AB, Upsalla, Sweden, equilibrated with 10 mM phosphate buffer (pH 6.6) before use. After washing the column with 10 mM phosphate buffer (pH 6.6) with a stepwise gradient of NaCl increasing from 0 M to 0.5 M was fed to the column, and fractions eluted by about 0.2 M NaCl were collected. The fractions were dialyzed against 10 mM phosphate buffer (pH 6.8) before applied to a column with "DEAE 5PW", a gel for ion-exchange chromatography, commercialized by TOSOH Corporation, Tokyo, Japan. To the column was fed 10 mM phosphate buffer (pH 6.8) with a linear gradient of NaCl increasing from 0 M to 0.5 M, and fractions eluted by about 0.2-0.3 M NaCl were collected.

While the obtained fractions were pooled and dialyzed against PBS, a gel for immunoaffinity chromatography with the monoclonal antibody were prepared according to the method disclosed in Japanese Patent Kokai No.231,598/96 by the same applicant of this invention. After the gel were charged into a plastic column and washed with PBS, the above dialyzed solution was applied to the column. To the column was fed 100 mM glycine-HCl buffer (pH 2.5), and the eluted fractions, which contained a polypeptide capable of inducing the production of IFN-γ by immunocompetent cells, were collected. After the collected fractions were dialyzed against sterilized distilled water and concentrated with a membrane filtration, the resultant was lyophilized to obtain a purified solid polypeptide in a yield of about 15 mg/l-culture.

#### Example for Reference

#### 35 Expression in Escherichia coli

As disclosed in Japanese Patent Kokai No.193,098/96, a transformant pKHuGF which was obtained by introducing a cDNA with the nucleotide sequence of SEQ ID NO:2 into Escherichia coli as a host, was inoculated into L-broth medium containing 50  $\mu$ g/ml ampicillin and cultured at 37°C for 18 hr under shaking conditions. The cells were collected by centrifuging the resulting culture, and then suspended in a mixture solution (pH 7.2) of 139 mM NaCl, 7 mM NaH<sub>2</sub>PO<sub>4</sub> and 3 mM Na<sub>2</sub>HPO<sub>4</sub>, followed by supersonicating to disrupt the cells. After the cell disruptants were centrifuged, the supernatant was subjected to purifying steps similarly as in Example 4-1 to obtain a purified solid polypeptide in a yield of about 5 mg/l-culture.

Comparing the yields of the polypeptides in Example for Reference and in Example 4-1 shows that the use of a transformant, which is formed by introducing a genomic DNA encoding the present polypeptide into a mammalian cell as a host, strongly elevates the yield of the polypeptide per culture.

#### Example 4-2

#### 50 Physicochemical property of polypeptide

#### Example 4-2(a)

#### **Biological activity**

55

45

Blood were collected from a healthy donor by using a syringe containing heparin, and then diluted with 2-fold volume of serum-free RPMI-1640 medium (pH 7.4). The blood was overlaid on ficoll, commercialized by Pharmacia LKB Biotechnology AB, Upsalla, Sweden, and centrifuged to obtain lymphocytes, which were then washed with RPMI-

1640 medium containing 10 v/v % bovine fetal serum before being suspended in a fresh preparation of the same medium to give a cell density of 5 x  $10^5$  cells/ml. 0.15 ml aliquots of the cell suspension was distributed into wells of micro plates with 96 wells.

To the wells with the cells were distributed 0.05 ml aliquots of solutions of the polypeptide in Example 4-1, diluted with RPMI-1640 medium (pH 7.4) containing 10 v/v % bovine fetal serum to give desired concentrations. 0.05 ml aliquots of fresh preparations of the same medium with 2.5  $\mu$ g/ml concanavalin A were further added to the wells, before culturing in a 5 v/v % CO<sub>2</sub> incubator at 37°C for 24 hr. After the cultivation, 0.1 ml of the culture supernatant was collected from each well and examined on IFN- $\gamma$  by usual enzyme immunoassay. In parallel, a systems as a control using the polypeptide in Reference for that in Example 4-1 or using no polypeptide was treated similarly as above. The results were in Table 1. IFN- $\gamma$  in Table 1 were expressed with international units (IU), calculated based on the IFN- $\gamma$  standard, Gg23-901-530, obtained from the International Institute of Health, USA

Table 1

Sample of polypeptide IFN-γ production (IU/ml)

Example 4-2(a) 3.4 x 10<sup>5</sup>

Example for Reference 1.7 x 10<sup>5</sup>

Table 1 indicates that the lymphocytes as immunocompetent cells produce IFN-γ by the action of the present polypeptide.

It is more remarkable that the polypeptide in Example 4-1 could induce IFN-γ production more than that in Example for Reference. Considering this and the difference in the yields of the polypeptides, described in Example for Reference, it can be presumed: Even if DNAs could be substantially equivalent in encoding the same amino acid sequence, not only the expressing efficiencies of the DNAs may differ, but the products expressed by them may significantly differ in their biological activities as a result of post-translational modifications by intracellular enzymes, depending on types of the DNAs and their hosts; (a) one type is used a transformant formed by introducing a DNA, which is a cDNA, into a microorganisms as a host, and (b) other type is used a transformant formed by introducing the present genomic DNA into a mammalian cell as a host.

#### Example 4-2(b)

10

15

20

25

35

40

45

50

#### Molecular weight

SDS-polyacrylamide gel electrophoresis of the polypeptide in Example 4-1 in the presence of 2 w/v % dithiothreitol as a reducing agent, according to the method reported by U. K. Laemli et al., in "Nature", Vol.227, pp.680-685 (1970), exhibited a main band of a protein capable of inducing IFN- $\gamma$  in a position corresponding to a molecular weight of about 18,000-19,500 daltons. The molecular weight makers used in the analysis were bovine serum albumin (67,000 daltons), ovalbumin (45,000 daltons), carbonic anhydrase (30,000 daltons), soy bean trypsin inhibitor (20,100 daltons) and  $\alpha$ -lactoalbumin (14,000 daltons).

#### Example 4-2(c)

#### N-Terminal amino acid sequence

Conventional analysis using "MODEL 473A", a protein sequencer commercialized by Perkin-Elmer Corp., Norwalk, USA, revealed that the polypeptide in Example 4-1 had the amino acid sequence of SEQ ID NO:15 in the N-terminal region.

Judging collectively from this result as well as the information that SDS-polyacrylamide gel electrophresis exhibited a main band in a position corresponding to a molecular weight of about 18,000-19,500 daltons, and that the molecular weight calculated from the amino acid sequence of SEQ ID NO:1 was 18,199 daltons, it can be concluded that the polypeptide in Example 4-1 has the amino acid sequence of SEQ ID NO:6.

As is described above, the present invention is made based on the identification of a genomic DNA encoding the polypeptide which induces the production of IFN- $\gamma$  by immunocompetent cells. The present genomic DNA efficiently express the present polypeptide when introduced into mammalian host cells. The polypeptide features higher biological activities than that obtained by the cDNA expression in *Escherichia coli*. Therefore, the present genomic DNA is useful for the recombinant DNA techniques to prepare the polypeptide capable of inducing IFN- $\gamma$  production by immunocompetent cells. The present genomic DNA is useful to gene therapy for diseases including viral diseases, bacterial-infectious diseases, malignant tumors and immunopathies.

Thus, the present invention is a significant invention which has a remarkable effect and gives a great contribution to this field.

While there has been described what is at present considered to be the preferred embodiments of the present invention, it will be understood the various modifications may be made therein, and it is intended to cover in the appended claims all such modifications as fall within the true spirits and scope of the invention.

#### SEQUENCE LISTING

(1) GENERAL	INFORMATION
-------------	-------------

(i) APPLICANT:

NAME: KABUSHIKI KAISHA HAYASHIBARA SEIBUTSU KAGAKU KENKYUJO

- (ii) TITLE OF INVENTION: GENOMIC DNA ENCODING A POLYPEPTIDE CAPABLE OF INDUCING THE PRODUCTION OF INTERFERON-  $\gamma$ 
  - (iii) NUMBER OF SEQUENCES:15
  - (iv) ADDRESS:
    - (A) ADDRESSEE: KABUSHIKI KAISHA HAYASHIBARA SEIBUTSU KAGAKU KENKYUJO
    - (B) STREET: 2-3, 1-CHOME, SHIMOISHII
    - (C) CITY:OKAYAMA
    - (E) COUNTRY: JAPAN
    - (F) POSTAL CODE (ZIP):700
- 20 (v) COMPUTER READABLE FORM:
  - (A) MEDIUM TYPE:Floppy disk
  - (B) COMPUTER: IBM PC compatible
  - (C) OPERATING SYSTEM: PC-DOS/MS-DOS
- 25 (vii) PRIOR APPLICATION DATA:
  - (A) APPLICATION NUMBER: JP 185,305/96
  - (B) FILING DATE: June 27, 1996
  - (2) INFORMATION FOR SEQ ID NO:1:
    - (i) SEQUENCE CHARACTERISTICS:
      - (A) LENGTH: 157 amino acids
      - (B) TYPE: amino acid
      - (D) TOPOLOGY: linear
    - (ii) MOLECULE TYPE: peptide
  - (xi) SEQUENCE DESCRIPTION: SEQ ID NO:1:

Tyr Phe Gly Lys Leu Glu Ser Lys Leu Ser Val Ile Arg Asn Leu Asn
1 10 15

Asp Gln Val Leu Phe Ile Asp Gln Gly Asn Arg Pro Leu Phe Glu Asp 20 25 30

Met Thr Asp Ser Asp Cys Arg Asp Asn Ala Pro Arg Thr Ile Phe Ile 35 40 45

Ile Ser Met Tyr Lys Asp Ser Gln Pro Arg Gly Met Ala Val Thr Ile
50 55 60

Ser Val Lys Cys Glu Lys Ile Ser Xaa Leu Ser Cys Glu Asn Lys Ile 65 70 75 80

Ile Ser Phe Lys Glu Met Asn Pro Pro Asp Asn Ile Lys Asp Thr Lys
85
90
95

Ser Asp Ile Ile Phe Phe Gln Arg Ser Val Pro Gly His Asp Asn Lys
100 105 110

Met Gln Phe Glu Ser Ser Ser Tyr Glu Gly Tyr Phe Leu Ala Cys Glu 115 120 125

Lys Glu Arg Asp Leu Phe Lys Leu Ile Leu Lys Lys Glu Asp Glu Leu
130
135
140
Gly Asp Arg Ser Ile Met Phe Thr Val Gln Asp Glu Asp

Gly Asp Arg Ser Ile Met Phe Thr Val Gln Asn Glu Asp 145 150 155

55

15

30

35

40

45

```
(3) INFORMATION FOR SEQ ID NO:2:
           (i) SEQUENCE CHARACTERISTICS:
                (A) LENGTH: 1120 base pairs
                (B) TYPE: nucleic acid
5
                (C) STRANDEDNESS: double
                (D) TOPOLOGY: linear
           (ii) MOLECULE TYPE: cDNA to mRNA
           (iii) HYPOTHETICAL: No
           (iv) ANTI-SENSE: No
10
           (vi)ORIGINAL SOURCE:
                (A) ORGANISM: human
                (F) TISSUE TYPE: liver
           (iX) FEATURE:
                (A) NAME/KEY:5'UTR
                (B) LOCATION: 1..177
15
                (C) IDENTIFICATION METHODS: E
                (A) NAME/KEY: leader peptide
                (B) LOCATION: 178..285
                (C) IDENTIFICATION METHODS:S
                (A) NAME/KEY: mat peptide
                (B) LOCATION: 286..756
20
                (C) IDENTIFICATION METHODS:S
                (A) NAME/KEY: 3'UTR
                (B) LOCATION: 757..1120
                (C) IDENTIFICATION METHODS: E
           (xi) SEQUENCE DESCRIPTION: SEO ID NO:2:
     GCCTGGACAG TCAGCAAGGA ATTGTCTCCC AGTGCATTTT GCCCTCCTGG CTGCCAACTC
     120
     CCTCAGACCT TCCAGATCGC TTCCTCTCGC AACAAACTAT TTGTCGCAGG AATAAAG
                                                                         177
     ATG GCT GCT GAA CCA GTA GAA GAC AAT TGC ATC AAC TTT GTG GCA ATG
     Met Ala Ala Glu Pro Val Glu Asp Asn Cys Ile Asn Phe Val Ala Met
30
                              -30
     AAA TTT ATT GAC AAT ACG CTT TAC TTT ATA GCT GAA GAT GAA AAC
     Lys Phe Ile Asp Asn Thr Leu Tyr Phe Ile Ala Glu Asp Asp Glu Asn
     -20
                          -15
                                              -10
     CTG GAA TCA GAT TAC TTT GGC AAG CTT GAA TCT AAA TTA TCA GTC ATA
     Leu Glu Ser Asp Tyr Phe Gly Lys Leu Glu Ser Lys Leu Ser Val Ile
35
     AGA AAT TTG AAT GAC CAA GTT CTC TTC ATT GAC CAA GGA AAT CGG CCT
                                                                         369
     Arg Asn Leu Asn Asp Gln Val Leu Phe Ile Asp Gln Gly Asn Arg Pro
                                 20
                                                     25
     CTA TTT GAA GAT ATG ACT GAT TCT GAC TGT AGA GAT AAT GCA CCC CGG
                                                                         417
     Leu Phe Glu Asp Met Thr Asp Ser Asp Cys Arg Asp Asn Ala Pro Arg
                             35
     ACC ATA TTT ATT ATA AGT ATG TAT AAA GAT AGC CAG CCT AGA GGT ATG
                                                                         465
     Thr Ile Phe Ile Ile Ser Met Tyr Lys Asp Ser Gln Pro Arg Gly Met
                                             55
     GCT GTA ACT ATC TCT GTG AAG TGT GAG AAA ATT TCA AYT CTC TCC TGT
                                                                         513
45
     Ala Val Thr Ile Ser Val Lys Cys Glu Lys Ile Ser Xaa Leu Ser Cys
                      65
                                          70
     GAG AAC AAA ATT ATT TCC TTT AAG GAA ATG AAT CCT CCT GAT AAC ATC
                                                                         561
     Glu Asn Lys Ile Ile Ser Phe Lys Glu Met Asn Pro Pro Asp Asn Ile
                                     85
                                                         90
     AAG GAT ACA AAA AGT GAC ATC ATA TTC TTT CAG AGA AGT GTC CCA GGA
                                                                         609
50
     Lys Asp Thr Lys Ser Asp Ile Ile Phe Phe Gln Arg Ser Val Pro Gly
             95
                                 100
                                                     105
     CAT GAT AAT AAG ATG CAA TTT GAA TCT TCA TCA TAC GAA GGA TAC TTT
                                                                         657
     His Asp Asn Lys Met Gln Phe Glu Ser Ser Ser Tyr Glu Gly Tyr Phe
         110
                             115
                                                 120
     CTA GCT TGT GAA AAA GAG AGA GAC CTT TTT AAA CTC ATT TTG AAA AAA
                                                                         705
```

5	125 GAG	GAT	GAA	TTG	GGG	Asp	AGA	TCT	ATA	ATG	135 TTC Phe	ACT	GTT	CAA	AAC	140 GAA	753
	GAC	TAG	CTAT	TAA .		TCAT	GC C	GGGC	GCAG'			CGCC	TGT	AATC	CCA		806
10	CCA CAT GTA AAC	CTTT ACAT GCCC GAGG TCCA	GGT ( TCA : TTG '	GAAA ATCC IGGT CAAA	CCTC. CAGC' GAGC	AT C' TA C' CG A	TCTA TCAA GATT	CTAA GAGG GCAC	A AA' C TG C AT'	TACT. AGGC. TGCG	AAAA AGGA CTCT	ATT GAA' AGC	AGCT( ICAC' ITGG(	GAG ( ITG ( GCA )	TGTA CACT ACAA	GCCTGA GTGACG CCGGAG CAGCAA TGAAAA	926 986 1046
15																	
	(4)					SEQ ARAC'											
,			() () ()	A) LEI B) TYI C) STI	NGTH PE : ni RANDI	:135 ucle: EDNE:	base ic ac SS:do	e pa: cid ouble	irs								
20			i) MOI i) OR:	LECUI IGINA A) ORG	LE T AL SO GANIS	GY:1: YPE:0 OURC! SM:hi TYP!	Genor E: ıman	nic I									
25		(i)	X) FEZ		_	EY : ex	ron										
		(x:	(1	B) LOC	CATIO ENTII	ON:1 FICAT ESCRI	139	METE			D:3:						
30						GAT Asp											47
		ATA				AAT Asn										AAT	95
35						GAT Asp											135
			204	30	<b>014</b>	p			35	501	nop	Cys	AL 9	40			
40	(5)		SEQU (2 (E (C	JENCE () LEN () TYP () STP	E CHA IGTH: PE:nu KANDE	SEQ ARACT :134 :clei	TERIS base c ac SS:do	TICS pai cid ouble	: rs								
45			L) MOI	ECUI	E T	3Y:li YPE:0	enon		NA								
70			(7	ORC T) TIS	SANIS	OURCE SM:hu TYPE	ıman	cent	:a								
50			( Z ( E	A) NAN B) LOC C) IDE	E/KI CATIO ENTII	EY:ex ON:1. FICAT ESCRI	.134	METH			D:4:						
55						ACC Thr 45											47

	CAG CCT AGA GGT ATG GCT GTA ACT ATC TCT GTG AAG TGT GAG AAA ATT Gln Pro Arg Gly Met Ala Val Thr Ile Ser Val Lys Cys Glu Lys Ile	95
5	TCA ACT CTC TCT TGT GAG AAC AAA ATT ATT TCC TTT AAG Ser Thr Leu Ser Cys Glu Asn Lys Ile Ile Ser Phe Lys 80 85	134
10	(6) INFORMATION FOR SEQ ID NO:5:    (i)SEQUENCE CHARACTERISTICS:    (A)LENGTH:87 base pairs    (B)TYPE:nucleic acid    (C)STRANDEDNESS:double	
15	<pre>(D)TOPOLOGY:linear (ii)MOLECULE TYPE:Genomic DNA (vi)ORIGINAL SOURCE:     (A)ORGANISM:human     (F)TISSUE TYPE:placenta (iX)FEATURE:</pre>	
20	(A) NAME/KEY: exon (B) LOCATION: 187 (C) IDENTIFICATION METHODS: S (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 5:	
	GAATAAAG ATG GCT GCT GAA CCA GTA GAA GAC AAT TGC ATC AAC TTT GTG Met Ala Ala Glu Pro Val Glu Asp Asn Cys Ile Asn Phe Val	50
25	-35 -30 -25  GCA ATG AAA TTT ATT GAC AAT ACG CTT TAC TTT ATA G  Ala Met Lys Phe Ile Asp Asn Thr Leu Tyr Phe Ile Ala -20 -15 -10	87
30	(7) INFORMATION FOR SEQ ID NO:6:     (i)SEQUENCE CHARACTERISTICS:     (A)LENGTH:12 base pairs     (B)TYPE:nucleic acid	
35	(C)STRANDEDNESS:double (D)TOPOLOGY:linear (ii)MOLECULE TYPE:Genomic DNA (vi)ORIGINAL SOURCE: (A)ORGANISM:human	
40	(xi) SEQUENCE DESCRIPTION: SEQ ID NO:6:	
45	CT GAA GAT GAT G Ala Glu Asp Asp Glu -10	12
50	(8) INFORMATION FOR SEQ ID NO:7:     (i)SEQUENCE CHARACTERISTICS:         (A)LENGTH:2167 base pairs         (B)TYPE:nucleic acid         (C)STRANDEDNESS:double         (D)TOPOLOGY:linear     (ii)MOLECULE TYPE:Genomic DNA	
55	(vi)ORIGINAL SOURCE:	

(A) ORGANISM: human

(9) INFORMATION FOR SEQ ID NO:8: (i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 1334 base pairs

```
(F) TISSUE TYPE:placenta
            (iX)FEATURE:
5
                  (A) NAME/KEY:exon + 3'UTR
                  (B) LOCATION: 1..2167
                   (C) IDENTIFICATION METHODS: E
             (xi) SEQUENCE DESCRIPTION: SEQ ID NO:7:
       GAA ATG AAT CCT CCT GAT AAC ATC AAG GAT ACA AAA AGT GAC ATC ATA
10
      Glu Met Asn Pro Pro Asp Asn Ile Lys Asp Thr Lys Ser Asp Ile Ile
                             90
                                                    95
      TTC TTT CAG AGA AGT GTC CCA GGA CAT GAT AAT AAG ATG CAA TTT GAA
                                                                                    96
       Phe Phe Gln Arg Ser Val Pro Gly His Asp Asn Lys Met Gln Phe Glu
                                               110
                                                                      115
      TCT TCA TCA TAC GAA GGA TAC TTT CTA GCT TGT GAA AAA GAG AGA GAC
15
      Ser Ser Ser Tyr Glu Gly Tyr Phe Leu Ala Cys Glu Lys Glu Arg Asp
                    120
                                           125
      CTT TTT AAA CTC ATT TTG AAA AAA GAG GAT GAA TTG GGG GAT AGA TCT
                                                                                   192
      Leu Phe Lys Leu Ile Leu Lys Lys Glu Asp Glu Leu Gly Asp Arg Ser
               135
                                      140
                                                             145
20
      ATA ATG TTC ACT GTT CAA AAC GAA GAC TAGCTAT TAAAATTTCA TGCCGGGCGC
      Ile Met Phe Thr Val Gln Asn Glu Asp
      AGTGGCTCAC GCCTGTAATC CCAGCCCTTT GGGAGGCTGA GGCGGGCAGA TCACCAGAGG
      TCAGGTGTTC AAGACCAGCC TGACCAACAT GGTGAAACCT CATCTCTACT AAAAATACAA
AAAATTAGCT GAGTGTAGTG ACCCATGCCC TCAATCCCAG CTACTCAAGA GGCTGAGGCA
GGAGAATCAC TTGCACTCCG GAGGTGGAGG TTGTGGTGAG CCGAGATTGC ACCATTGCGC
25
                                                                                   486
      TCTAGCCTGG GCAACACAG CAAAACTCCA TCTCAAAAAA TAAAATAAAT AAATAAACAA
      ATAAAAAATT CATAATGTGA ACTGTCTGAA TTTTTATGTT TAGAAAGATT ATGAGATTAT
      TAGTCTATAA TTGTAATGGT GAAATAAAAT AAATACCAGT CTTGAAAAAC ATCATTAAGA
                                                                                  666
      AATGAATGAA CTTTCACAAA AGCAAACAAA CAGACTTTCC CTTATTTAAG TGAATAAAAT 726
      AAAATAAAAT AAAATAATGT TTAAAAAATT CATAGTTTGA AAACATTCTA CATTGTTAAT 786
      TGGCATATTA ATTATACTTA ATATAATTAT TTTTAAATCT TTTGGGTTAT TAGTCCTAAT 846
      GACAAAAGAT ATTGATATTT GAACTTTCTA ATTTTTAAGA ATATCGTTAA ACCATCAATA 906
      TTTTATAAG GAGGCCACTT CACTTGACAA ATTTCTGAAT TTCCTCCAAA GTCAGTATAT .966
TTTTAAAAATT CAGTTTGATC CTGAATCCAG CAATATATAA AAGGGATTAT ATACTCTGGC 1026
      CAACTGACAT TCATCCTAGG AATGCAAAGA TGGTTTAATA TCCTAAAATC AATTAACATA 1086
35
      ACATACTATA TTAATAAAGT ATCAAAACAG TATTCTCATC TTTTTTCTT TTTTCACAAT 1146
      TCCTTGGTTA CACTATCATC TCAATAGATG CAGAAAAAGC ATTTGACAAA ATCCAATTCA 1206
      TAATAAAAAT TCTCAAACTT GAAAGAGAAC ATCATAAAGG CATCTATGAA AAACCTACAG 1266
      CTAATATCAT ACTTAACGAT GAAAAACTGA ATTATTTTAC CCTAAGATCA AGAATAATGC 1326
      AAGCATGTCA GCTCTTGCAA CTTCTATTCA ACATTGTACT GGAGGTTCTA GCCAGAGCAA 1386
      CCATACAATA AATAAAAATA AAAGGCACCC AGATTAGAAA GGAAGTCTTT ATTTGCAGAC 1446
      AACATGGTTC TTTATGCAGA AAACCGTCAG GAATACACAC ACATGTTAGA ACTAATAAGT 1506
      TCAGCAAGGT TGCAGGTTGC AATATCAATA TGCAAAAATA CATTGAAGGC TGGGCTCAGT 1566
      GGAGATGGCA TGTACCTTTC GTCCCAGCTA CTTGGGAGGC TGAGGTAGGA GGATCACTTG 1626
      AGGTGAGGAG TTTGAGGCTA TAGTGCAATG TGATCTTGCC TGTGAATAGC CACTGCACTC 1686
      GAGCCTAGGC AACAAAGTGA GACCCCGTCT CCAAAAAAAA AAATGGTATA TTGGTATTTC 1746
      TGTATATGAA CAATGAATGA TCTGAAAACA AGAAAATTCC ATTCACGATG GTATTAAAAA 1806
AATAAAATAC AAATAAATTT AGCAAAATAA TTATAAAACT TGTACATCGA AAATTTCAAA 1866
GCACTCTGAG GGAAATTAAA GATGATCTAA ATAATTGGAG AGACACTCTA TGATCACTGA 1926
45
      TTGGAAAATT CATTCAATAT TGTTAAGATA ACAATTGTCC CCAAATTGAT GCATGCATTC 1986
      AATTTAGTCT TCATCAAAAT TCCAGCAGGG TTTTTGCAGA AATTGACAAG CTGTACCCAA 2046
      AATGTATATG GAAATGAAAA GACCCAGAAG AGCAAATAAT TTTTTAAAAA CAAAGTTGGA 2106
      AAACTTTTAC TTCCTAATTT TAAAACTTAC TATAAACCTA AAGTTATCAA GACCATTTAG 2166
50
```

```
(B) TYPE: nucleic acid
                   (C) STRANDEDNESS: double
                   (D) TOPOLOGY: linear
             (ii) MOLECULE TYPE: Genomic DNA
5
             (vi)ORIGINAL SOURCE:
                   (A) ORGANISM: human
                   (F) TISSUE TYPE: placenta
             (iX) FEATURE:
                   (A) NAME/KEY: intron
10
                   (B) LOCATION: 1.. 1334
                   (C) IDENTIFICATION METHODS: E
             (xi) SEQUENCE DESCRIPTION: SEQ ID NO:8:
      GTATTTTTT TAATTCGCAA ACATAGAAAT GACTAGCTAC TTCTTCCCAT TCTGTTTTAC
      TGCTTACATT GTTCCGTGCT AGTCCCAATC CTCAGATGAA AAGTCACAGG AGTGACAATA
                                                                                    120
15
      ATTTCACTTA CAGGAAACTT TATAAGGCAT CCACGTTTTT TAGTTGGGGT AAAAAATTGG
                                                                                    180
      ATACAATAAG ACATTGCTAG GGGTCATGCC TCTCTGAGCC TGCCTTTGAA TCACCAATCC
      CTTTATTGTG ATTGCATTAA CTGTTTAAAA CCTCTATAGT TGGATGCTTA ATCCCTGCTT
      GTTACAGCTG AAAATGCTGA TAGTTTACCA GGTGTGGTGG CATCTATCTG TAATCCTAGC TACTTGGGAG GCTCAAGCAG GAGGATTGCT TGAGGCCAGG ACTTTGAGGC TGTAGTACAC
                                                                                    420
      TGTGATCGTA CCTGTGAATA GCCACTGCAC TCCAGCCTGG GTGATATACA GACCTTGTCT
20
      CTAAAATTAA AAAAAAAAA AAAAAAAACC TTAGGAAAGG AAATTGATCA AGTCTACTGT
      GCCTTCCAAA ACATGAATTC CAAATATCAA AGTTAGGCTG AGTTGAAGCA GTGAATGTGC
                                                                                    600
      ATTCTTTAAA AATACTGAAT ACTTACCTTA ACATATATTT TAAATATTTT ATTTAGCATT
      25
      GGCATGCACC ACTTACACCC GGCTAATTTT TGTATTTTTA GTAGAGCTGG GGTTTCACCA
TGTTGGCCAG GCTGGTCTCA AACCCCTAAC CTCATACACCC
                                                                                    900
      TGTTGGCCAG GCTGGTCTCA AACCCCTAAC CTCAAGTGAT CTGCCTGCCT CAGCCTCCCA 1020
      AACAAACAAA CAACCCCACA GTTTAATATG TGTTACAACA CACATGCTGC AACTTTTATG 1080
      AGTATTTTAA TGATATAGAT TATAAAAGGT TGTTTTTAAC TTTTAAATGC TGGGATTACA 1140
      GGCATGAGCC ACTGTGCCAG GCCTGAACTG TGTTTTTAAA AATGTCTGAC CAGCTGTACA 1200
TAGTCTCCTG CAGACTGGCC AAGTCTCAAA GTGGGAACAG GTGTATTAAG GACTATCCTT 1260
TGGTTAAATT TCCGCAAATG TTCCTGTGCA AGAATTCTTC TAACTAGAGT TCTCATTTAT 1320
30
      TATATTTATT TCAG
35
      (10) INFORMATION FOR SEQ ID NO:9:
            (i) SEQUENCE CHARACTERISTICS:
                  (A) LENGTH: 4773 base pairs
                  (B) TYPE: nucleic acid
                  (C) STRANDEDNESS: double
                  (D) TOPOLOGY: linear
40
            (ii) MOLECULE TYPE: Genomic DNA
            (vi)ORIGINAL SOURCE:
                  (A) ORGANISM: human
                  (F) TISSUE TYPE: placenta
            (iX) FEATURE:
                  (A) NAME/KEY: intron
45
                  (B) LOCATION: 1..4773
                  (C) IDENTIFICATION METHODS: E
            (xi) SEQUENCE DESCRIPTION: SEQ ID NO:9:
     GTAAGACTGA GCCTTACTTT GTTTTCAATC ATGTTAATAT AATCAATATA ATTAGAAATA
50
     TAACATTATT TCTAATGTTA ATATAAGTAA TGTAATTAGA AAACTCAAAT ATCCTCAGAC
                                                                                   120
     CAACCTTTTG TCTAGAACAG AAATAACAAG AAGCAGAGAA CCATTAAAGT GAATACTTAC TAAAAATTAT CAAACTCTTT ACCTATTGTG ATAATGATGG TTTTTCTGAG CCTGTCACAG
                                                                                   180
                                                                                   240
     GGGAAGAGGA GATACAACAC TTGTTTTATG ACCTGCATCT CCTGAACAAT CAGTCTTTAT
                                                                                    300
     ACAAATAATA ATGTAGAATA CATATGTGAG TTATACATTT AAGAATAACA TGTGACTTTC
                                                                                   360
     CAGAATGAGT TCTGCTATGA AGAATGAAGC TAATTATCCT TCTATATTTC TACACCTTTG 420
55
```

```
TAAATTATGA TAATATTTTA ATCCCTAGTT GTTTTGTTGC TGATCCTTAG CCTAAGTCTT 480
         AGACACAAGC TTCAGCTTCC AGTTGATGTA TGTTATTTTT AATGTTAATC TAATTGAATA 540
         AAAGTTATGA GATCAGCTGT AAAAGTAATG CTATAATTAT CTTCAAGCCA GGTATAAAGT 600
ATTTCTGGCC TCTACTTTTT CTCTATTATT CTCCATTATT ATTCTCTATT ATTTTTCTCT 660
 5
         ATTTCCTCCA TTATTGTTAG ATAAACCACA ATTAACTATA GCTACAGACT GAGCCAGTAA 720
         GAGTAGCCAG GGATGCTTAC AAATTGGCAA TGCTTCAGAG GAGAATTCCA TGTCATGAAG 780
         ACTCTTTTTG AGTGGAGATT TGCCAATAAA TATCCGCTTT CATGCCCACC CAGTCCCCAC 840
         TGAAAGACAG TTAGGATATG ACCTTAGTGA AGGTACCAAG GGGCAACTTG GTAGGGAGAA 900
         AAAAGCCACT CTAAAATATA ATCCAAGTAA GAACAGTGCA TATGCAACAG ATACAGCCCC 950
CAGACAAATC CCTCAGCTAT CTCCCTCCAA CCAGAGTGCC ACCCCTTCAG GTGACAATTT 1020
10
         GGAGTCCCCA TTCTAGACCT GACAGGCAGC TTAGTTATCA AAATAGCATA AGAGGCCTGG 1080
         GATGGAAGGG TAGGGTGGAA AGGGTTAAGC ATGCTGTTAC TGAACAACAT AATTAGAAGG 1140
         GAAGGAGATG GCCAAGCTCA AGCTATGTGG GATAGAGGAA AACTCAGCTG CAGAGGCAGA 1200
         TTCAGAAACT GGGATAAGTC CGAACCTACA GGTGGATTCT TGTTGAGGGA GACTGGTGAA 1260
         AATGTTAAGA AGATGGAAAT AATGCTTGGC ACTTAGTAGG AACTGGGCAA ATCCATATTT 1320
         GGGGGAGCCT GAAGTTTATT CAATTTTGAT GGCCCTTTTA AATAAAAAGA ATGTGGCTGG 1380
GCGTGGTGGC TCACACCTGT AATCCCAGCA CTTTGGGAGG CCGAGGGGGG CGGATCACCT 1440
GAAGTCAGGA GTTCAAGACC AGCCTGACCA ACATGGAGAA ACCCCATCTC TACTAAAAAT 1500
ACAAAATTAG CTGGGGTGG TGGCATATGC CTGTAATCCC AGCTACTCTG GAGGCTGAGG 1560
15
         CAGGAGAATC TTTTGAACCC GGGAGGCAGA GGTTGCGATG AGCCTAGATC GTGCCATTGC 1620
        20
         GCAGTGGTTA TAAAAGTGGC CTAGGTTCTA GATAATAAGA TACAACAGGC CAGGCACAGT 1980
         GGCTCATGCC TATAATCCCA GCACTTTGGG AGGGCAAGGC GAGTGTCTCA CTTGAGATCA 2040
        GGAGTTCAAG ACCAGCCTGG CCAGCATGGC GATACTCTGT CTCTACTAAA AAAAATACAA 2100
AAATTAGCCA GGCATGGTGG CATGCACCTG TAATCCCAGC TACTCGTGAG CCTGAGGCAG 2160
AAGAATCGCT TGAAACCAGG AGGTGTAGGC TGCAGTGAGC TGAGATCGCA CCACTGCACT 2220
CCAGCCTGGG CGACAGAATG AGACTTTGTC TCAAAAAAAG AAAAAGATAC AACAGGCTAC 2280
25
        CCTTATGTGC TCACCTTTCA CTGTTGATTA CTAGCTATAA AGTTCTTTGG 2340
        TCAAGAACCT TGACAACACT AAGAGGGATT TGCTTTGAGA GGTTACTGTC AGAGTCTGTT 2400
30
        TCATATATAT ACATATACAT GTATATATGT ATCTATATCC AGGCTTGGCC AGGGTTCCCT 2460
        CAGACTTTCC AGTGCACTTG GGAGATGTTA GGTCAATATC AACTTTCCCT GGATTCAGAT 2520
        TCAACCCCTT CTGATGTAAA AAAAAAAAA AAAAAGAAAG AAATCCCTTT CCCCTTGGAG 2580 CACTCAAGTT TCACCAGGTG GGGCTTTCCA AGTTGGGGGT TCTCCAAGGT CATTGGGATT 2640 GCTTTCACAT CCATTTGCTA TGTACCTTCC CTATGATGGC TGGGAGTGGT CAACATCAAA 2700
        ACTAGGAAAG CTACTGCCCA AGGATGTCCT TACCTCTATT CTGAAATGTG CAATAAGTGT 2760
35
        GATTAAAGAG ATTGCCTGTT CTACCTATCC ACACTCTCGC TTTCAACTGT AACTTTCTTT 2820
        TTTTCTTTTT TTCTTTTTTT CTTTTTTTT GAAACGGAGT CTCGCTCTGT CGCCCAGGCT 2880 AGAGTGCAGT GGCACGATCT CAGCTCACTG CAAGCTCTGC CTCCCGGGTT CACGCCATTC 2940 TCCTGCCTCA CCCTCCCAAG CAGCTGGGAC TACAGGCGCC TGCCACCATG CCCAGCTAAT 3000 TTTTTGTATT TTTAGTAGAG ACGGGGTTTC ACCGTGTTAG CCAGGATGGT CTCGATCTCC 3060
        TGAACTTGTG ATCCGCCCGC CTCAGCCTCC CAAAGTGCTG GGATTACAGG CGTGAGCCAT 3120
40
        CGCACCCGGC TCAACTGTAA CTTTCTATAC TGGTTCATCT TCCCCTGTAA TGTTACTAGA 3180
        GCTTTTGAAG TTTTGGCTAT GGATTATTTC TCATTTATAC ATTAGATTTC AGATTAGTTC 3240
        CAAATTGATG CCCACAGCTT AGGGTCTCTT CCTAAATTGT ATATTGTAGA CAGCTGCAGA 3300
AGTGGGTGCC AATAGGGGAA CTAGTTTATA CTTTCATCAA CTTAGGACCC ACACTTGTTG 3360
ATAAAGAACA AAGGTCAAGA GTTATGACTA CTGATTCCAC AACTGATTGA GAAGTTGGAG 3420
        ATAACCCCGT GACCTCTGCC ATCCAGAGTC TTTCAGGCAT CTTTGAAGGA TGAAGAAATG 3480
45
        CTATTTTAAT TTTGGAGGTT TCTCTATCAG TGCTTAGGAT CATGGGAATC TGTGCTGCCA 3540
        TGAGGCCAAA ATTAAGTCCA AAACATCTAC TGGTTCCAGG ATTAACATGG AAGAACCTTA 3600
        GGTGGTGCCC ACATGTTCTG ATCCATCCTG CAAAATAGAC ATGCTGCACT AACAGGAAAA 3660
GTGCAGGCAG CACTACCAGT TGGATAACCT GCAAGATTAT AGTTTCAAGT AATCTAACCA 3720
        TTTCTCACAA GGCCCTATTC TGTGACTGAA ACATACAAGA ATCTGCATTT GGCCTTCTAA 3780
50
        GGCAGGGCCC AGCCAAGGAG ACCATATTCA GGACAGAAAT TCAAGACTAC TATGGAACTG 3840
        GAGTGCTTGG CAGGGAAGAC AGAGTCAAGG ACTGCCAACT GAGCCAATAC AGCAGGCTTA 3900
        CACAGGAACC CAGGGCCTAG CCCTACAACA ATTATTGGGT CTATTCACTG TAAGTTTTAA 3960
        TTTCAGGCTC CACTGAAAGA GTAAGCTAAG ATTCCTGGCA CTTTCTGTCT CTCTCACAGT 4020
TGGCTCAGAA ATGAGAACTG GTCAGGCCAG GCATGGTGGC TTACACCTGG AATCCCAGCA 4080
```

```
CTTTGGGAGG CCGAAGTGGG AGGGTCACTT GAGGCCAGGA GTTCAGGACC AGCTTAGGCA 4140
         ACAAAGTGAG ATACCCCCTG ACCCCTTCTC TACAAAAATA AATTTTAAAA ATTAGCCAAA 4200
         TGTGGTGGTG TATACTTACA GTCCCAGCTA CTCAGGAGGC TGAGGCAGGG GGATTGCTTG 4260
        AGCCCAGGAA TTCAAGGCTG CAGTGAGCTA TGATTTCACC ACTGCACTTC TGGCTGGGCA 4320
5
        ACAGAGCGAG ACCCTGTCTC AAAGCAAAAA GAAAAAGAAA CTAGAACTAG CCTAAGTTTG 4380
TGGGAGGAGG TCATCATCGT CTTTAGCCGT GAATGGTTAT TATAGAGGAC AGAAATTGAC 4440
         ATTAGCCCAA AAAGCTTGTG GTCTTTGCTG GAACTCTACT TAATCTTGAG CAAATGTGGA 4500
         CACCACTCAA TGGGAGAGGA GAGAAGTAAG CTGTTTGATG TATAGGGGAA AACTAGAGGC 4560
         CTGGAACTGA ATATGCATCC CATGACAGGG AGAATAGGAG ATTCGGAGTT AAGAAGGAGA 4620
        GGAGGTCAGT ACTGCTGTTC AGAGATTTTT TTTATGTAAC TCTTGAGAAG CAAAACTACT 4680
10
        TTTGTTCTGT TTGGTAATAT ACTTCAAAAC AAACTTCATA TATTCAAATT GTTCATGTCC 4740
        TGAAATAATT AGGTAATGTT TTTTTCTCTA TAG
         (11) INFORMATION FOR SEQ ID NO:10:
15
               (i) SEQUENCE CHARACTERISTICS:
                     (A) LENGTH: 8835 base pairs
                     (B) TYPE: nucleic acid
                     (C) STRANDEDNESS: double
                     (D) TOPOLOGY: linear
               (ii) MOLECULE TYPE: Genomic DNA
20
               (vi)ORIGINAL SOURCE:
                     (A) ORGANISM: human
                     (F) TISSUE TYPE: placenta
              (iX) FEATURE:
                     (A) NAME/KEY: intron
                     (B) LOCATION: 1..8835
25
                     (C) IDENTIFICATION METHODS: E
               (xi) SEQUENCE DESCRIPTION: SEQ ID NO:10:
        GTAAGAATA TCATTCCTCT TTATTTGGAA AGTCAGCCAT GGCAATTAGA GGTAAATAAG
        CTAGAAAGCA ATTGAGAGGA ATATAAACCA TCTAGCATCA CTACGATGAG CAGTCAGTAT
                                                                                            120
        CAACATAAGA AATATAAGCA AAGTCAGAGT AGAATTTTTT TCTTTTATCA GATATGGGAG
                                                                                            180
30
        AGTATCACTT TAGAGGAGAG GTTCTCAAAC TTTTTGCTCT CATGTTCCCT TTACACTAAG
                                                                                            240
        CACATCACAT GTTAGCATAA GTAACATTTT TAATTAAAAA TAACTATGTA CTTTTTTAAC
                                                                                            300
        AACAAAAAA AGCATAAAGA GTGACACTTT TTTATTTTTA CAAGTGTTTT AACTGGTTTA
                                                                                            360
       ATAGAAGCCA TATAGATCTG CTGGATTCTC ATCTGCTTTG CATTCAGACT ACTGCAATAT TGCACAGAAT GCAGCCTCTG GTAAACTCTG TTGTACACTC ATGAGAGAAT GGGTGAAAAA
                                                                                            420
                                                                                            480
        GACAAATTAC GTCTTAGAAT TATTAGAAAT AGCTTTCACT TTAGGAACTC CCTGAGAATT
35
                                                                                            540
        GCTGCTTTAG AGTGGTAAGA TAAATAAGCT TCTCTTTAAA CGGAATCTCA AGACAGAATC
                                                                                            600
        AGTTACATTA AAAGCAAACA AAAAATTTGC CCATGGTTAG TCATCTTGTG AAATCTGCCA
                                                                                            660
        CACCTTTGGA CTGGGCTACA ATTGGATAAT ATAGCATTCC CCGAGATAAT TTTCTCTCAC
       AATTAAGGAA AGGGCTGAAT AAATATCTCT GTTTGAAGTT GAATAACAAA AATTAGGACC CCCTAAATTT TAGGGCTCCT GAAATTCGTC TTTTTGCCTA TATTCAGCTA CTTTACGTTC TATTAAATCT TCTTTCAGGC CAGGTGCACT AGCTCATGCC TAGAATCTCA GGCAGGCCTG
                                                                                            780
                                                                                            840
40
                                                                                            900
       AGCCCAGGAA TTTGAGACCA GCCAGGGCAA CACAGTCTCT ACAAAAAAAT AAAAAATTAC
                                                                                            960
       CTGGGTGTGT TGGTGCATGC CTGTAGAACT ACTCAGGATG CTGAGGACTG CTTGAGCCCA 1020
       GGATAGCCAA ATCTGTGGTG AGTTCAGCCA CTAAACAGAG CGAGACTTTC TCAAAAAAAC 1080
       AAACAAAAAA ACAAACAAAC TTCCTTCAAA ATAACTTTTT ATCTGCAATG TTTTCCTATT 1140
GCCTGTGAGA TTAAATTTAC TCTTTTACCT GATTTCCAAA GCCCTCCATA ATCTAATCCG 1200
ACTTTACCTT GTGTTCACTG CAAAATAGCA GGACTGTTCC ACTACAATCC AAAAATCACA 1260
45
       GGTTGGGTGC AGTGGCTCAC TCCTGTAATC CCAACACTTT GGAAGGCCAA GGCAGGTGGA 1320
       TTGCTTCAGC TCAGGAGTTC AAGACCAGCC TGGGCAACAT GGCAAAAACC CTGTCTCTCC 1380
       AAAACATACA AAAATTAGCC AGATGTGGTA GTATGTGCCT GTAGTCCCAA CTACTCAAAA 1440
GGCTAAGGCA AGAGGATCAC TTGAGCCCAG GAGGTCAAGG CTACAGTGAG CCATGTTTAC 1500
TGTGTCACTG CACTCCAGCC TGGGTGATAG AGCAAGACCA TGTCTCAAAA AAAAAAAAA 1560
       GAAAAGAAAA GAAAAAAACA TCGCTCTATT CAGTTCACCC CCACCACAAC ATTGTTTTGA 1620
TTATCACATA AATGCTGGTC CATTGCCTTC TCTATCTATT CAAATCTTTA AGCATTCTTT 1680
50
       GAGATTCAAC TCAATTCTCC TTTTCAAACT AGGCCATTTA AACTACATCA GTTCCATTTT
                                                                                          1740
       GATTTTCTTG CTTTGAGTCT ACAGACTCAA AAACAAAAAC TTAAAAACTT ATTTTTTAAG 1800
       TTTTCTGCTA CTCTCACTTC TTCAACACTC ACATACACGC ATTCATAATA AGATGGCAGA 1860
```

19

```
ATGTTCAAGG ATAAAATGAT TTATAGAACT GAAAAGTTAG GTTTTGATCT ŢGTTGCTGTC 1920
           AAGATGACTA CCTACCTGAT CTCAGGTAAT TAATTATGTA GCATGCTCCC TCATTTCATC 1980
CCATACCTAT TCAACAGGAT TGGAATTCCA CAGCAAGGAT AAACATAATC ATAGTTGCTT 2040
TTCAAGTTCA AGGCATTTA ACTTTTAATC TAGTAGTATG TTTGTTGTTG TTGTTGTTGT 2100
 5
           TTGAGATGGA GCCCTGCTGT GTCACCCAGG CTGGAGTGCA GTGGCACGAA CTCGGCTCAC 2160
           TGCAACCTCT GCCTCATGGG TTCAATCAGT TATTCTGCCT CAGTGTCCCA AGTAGCTGGG 2220
           ACTACAAGGC ACATGCCACC ATGCCTGGCT AATTTTTGTA TTTTTAGTAG AAACAGGGCT 2280
           TCACCATGTT GGCCAGGCTG GTCTCGAACT CCTGACCTCA AGTGATCCAG CCGCCTCGGC 2340
           CTCCCAAAGT GCTGGGATTA CAGGCATAAG CCACCGTGCC CAGCCTAATA GTATGTTTTT 2400
           AAACTCTTAG TGGCTTAACA ATGCTGGTTG TATAATAAAT ATGCCATAAA TATTTACTGT 2460 CTTAGAATTA TGAAGAAGTG GTTACTAGGC CGTTTGCCAC ATATCAATGG TTCTCTCCTT 2520 ACAGCTTTAA TTAGAGTCTA GAATTGCAGG TTGGTAGAGC TGGAACAGAC CTTAAAGATT 2580 GACTAGCCAA CTTCCTTGTC CAAATGAGGG AACTGAGACC CTTAAAATTA AGTGACTTGC 2640
10
           CCCAGACAAA ACTGGAACTC ATGTGTCCTA ATTTCCATCA TGAAATTCTA CCATTCACTA 2700
           GCCTCTGGCT AGTTGTCAAA GTATTGCATA ACTAAATTTT TATGTCTGTT TTAAAGAACA 2760
15
          AATTGTCACT GCTTACTCCT GGGAGGGTCT TTCTGAGGTG GTTTATAACT CTTAAAAAA 2820
AAAAAGTCAG TAGTCTGAGA ATTTTAGACG AAATAGTCAA AGCATTTTTA TCCAATGGAT 2880
CTATAATTTT CATAGATTAG AGTTAAATCA AAGAAACACG GATGAGAAAG GAAGAGGAAA 2940
ATTGAGGAGA GGAGGAATGG GGATGAGAAC ACACTACTTG TAATCAGTCA TAGATGTACT 3000
           GAGAACTAAC AAGAAGAATT GTAAGAAAAT AAGAATGAAG AATTCAAAAT CAACACATGA 3060
           AATAAAAAGA AACTACTAGG GAAAAATGGA GAAGACATTA GAAAAATTAT TCTATTTTTA 3120
20
           AAATTCTGTT TTCAGGCTTC CCTCCTGTTC TTCCTCCTTC TCATTGGTTT TCAGGTGGAG 3180
          GGAAAGTTTA AGATGGAAAA AATATATATA TTCTACACAT CCCTTTCTAC GCTGTTGTCA 3240
TGGCAACAAG GTTTATCATA GCAAACTTTT ATTCATACAA CATTTATTGA GTTCTTACTG 3300
           TGTGGTAAGC TCTTTCCAGG TGTTGAAAAT TCAGGGGAAA AAAGACAACT CATTGTCTTA 3360
           AAACTCAGAT GAAAGCTGAA CAGACCTATT TTTAATCAAA GTAATCTCAA TTTAGGGTAG 3420
          TAAGAGCTAT TTAAGAAGCA TGAACAGGTG TGAAGGAGGT AGGACTCTGA GGAGAGAATA 3480
          GTTAGCTAGG AATGAAAGAG CAGAGAAGTT TTCCTAGAGG AACTATTAAA GCTGGGAGTT 3540
          ACGGGATGAA AGATGAGGCA GGGTTTGCAG GCAAAAAAAA AAAAAAGGCA GGGGAAGGGG 3600
AAGTTCTGGC CTGGCAGAGA GAATAACTGT GGCAACAATG GAGGAGAGTC TGGAAGCAAG 3660
AAAACCAAGT AGAAGAGTAT TAAAATAGAA GATGCCAGGG GTAATGAGGG CTTGATTTAA 3720
          AACAGTGCTG TTGGAGATGG AGAGGAGATA CCAAATTCTG GAGACATTTC TGAGTTAGAA 3780
          CCTACAGTAT TTATCAGACA AGGGAAAGAT TAGACAAAGG AGTTAAGAAT GACTCCCAGG 3840
30
          TTTCAGTTTG GGGCAGGTAA CTAGGACATG TTTTGAAAAG TAATGTATTG GATCTCTTAC 3900
          CATTGGAACT ATGTATGTGG AGCCAAATTA AAATTTGTAC ATGTATATAA CTCTCCCCC 3960 ACCACCACTA ACTACTTCCC TAACTCTCTA CTTTGTAGCC AGACTTCCTA AAAGAATAGT 4020 TTGTAGTCAC TGTCTTTACT TTTCCCCTCC CATTCTGTCC TAGATATTTG TCCACCTACC 4080 ATCTGCTGCC TCCACTTTAC CCAAACTGTT CTACGGTTGC CCAAAACTTC CTAATTGCCA 4140
35
          AATTCAATGA ACAAGTTTAA GCTTATATGT AAATTAGGAG CTCTACAGTT TGATTTCGAG 4200
          CAGCCCCTCC TGAAACCCTT TCTCTTTCGA CTTCTGTGAC ACATCTCAGA TTTACAAAAC 4260
          TGAACTAATT ATTTTACACT TGAGCTGTAT TTTCGTTCTT CTTTCTTGAT GAATGAGGTA 4320
ACCACTCAAC AAATTGCCCA AGCCAAAAAC TACGAAGTCA TCCTCAGTTC CTCCTTCTTC 4380
TGTTTGACCC ACAACAGATC AGCTGAGAAA TCCCGCTGTT TAGTATCTCT TGAATTCATT 4440
ACCTTAATTT ATAGCCTCAT CAACTCTTAA TTGTTAAAAT TACTTCAGTA GTTGTTGTCT 4500
40
          GACCTCTGTC CAATCTTGTT CAATCAGGTC CATTCTTTTG TTCTTGGTGG TGGTGGTGGT 4560
          GTTGACAGAG TTTCGCTTTT GCTGCCCAGG CTGAAGTGCA GTGGAGCACT TCACTGCAAC 4620
          CACAGCCTCC TGGGTTTAAG CAGTTCACCC TCCCGAGTAG CTGGGACTAC AGGTATGTGC 4680 CACCACACCC AGCTAATTTT GTGTTTTCAG TAGAGACAGG GTTTCACCAT GTTGGTCAGG 4740 CTGGTCTCAA ACTCCTGACC TCAAGCAATC CACCCACCTC AGCCTCCCAA AGTGCTGGGA 4800 TTACAGGCAT GAGCCACTGC ACACGGACCA GATCCATTGT TTATGTTGCT TCTAGAGTGA 4860
45
          GTTTTTAAAA CACAAATTTG ACCATATCTT TCTCCAATTT AAGTCAGTAT TTTTTTTTTC 4920
          AGGAAAAAC AGTTCAAACT CTTTAGTCTG CTTACACAAG GCCTTTGTAG TCTGACTCTT 4980
          CTTTCCAAGC TTTCATCAAA GTATACTGCA AGTTACATTT TATGTGAATT GAATTAGGCA 5040
         ACGGTATAAA AATTATAGTT TATATGGGCA AAATGGAAAT AATGTTAACT CTTCCAAATA 5100 GTTTATCTAG AATGACATAA TTTCAAAGCT GTCAGGTCAA ATGAGTTATA AACTGTTAAC 5160 ACTATTGCCA CATGCAAGTG TCTCTTATAC TTGGTAGAAT TATCTGCTTC CATGTCATTA 5220 TTATGTAAAT TAGACTTTAA ATAACTCAGA AGTTCTTCAG ACATACAGGT TATTATTGTG 5280
50
          CTTTTTAAAC ATAATTTTAA ATAATTTTAT ATATGATAAT GTTATCCAAG TGCTAAGGGA 5340
          TGTATTGTTA CTGCTGTGCA AAAAAAAAA AAAAAAAAAC TCCAAATAAA TATGTTGAAA 5400
          CCAAGTTTAT ATGCAAGAAA ACAATATTAA AAAGGCCAAA GTACCACCAT AATAGGCTGT 5460
         GTGGAGACGG CAGGCTACAA AACACTAGTA ATAATGCTGA GAAAGTTGAA AAAAGAAAGA 5520
55
```

```
AAGCAACAAT ATGCTTTGGT TGTTGTAGGT TTATGTACTC CAAGAATATC TCCTCTCAAA 5580 CTTTTACGTT TTTTCCAAAG AAAAGTTAAC TTTGGCTGGG CGCAGTGGCT CTTGCCTGTA 5640 GTCCCAGCCT TTGGGAGGCC AAGGCGGGCA GATCACCTGA GGTCAGGAGT TTGAGACCAG 5700
        CCTGACCAAA AATGGAGAAA CCCGCCCCC TCACTACTAA AAGAATACAA AATTAGGCCG 5760
5
        GGCACAGTGG CTTACCCCTG TGATCCCAGC ACTTTGGGAG GCCGAAGCAG GAAGATCACC 5820
        TGAGGTCAGG AGTTCGAGAC CAGCCATGGA GAAACCCGTC TCTACTAAAA ATACAAATT 5880
        AGCCGGGCGT GGTGGTGCAT GACTGTAATC CCAGCTACTC AGGAGGCTAA GGCAGAGAAT 5940
        CACTTGAACC CAGGCAGTGG AGGTTGCAGT GAGCCGAGAT CGTGCCATTG CACTCCAGCC 6000
        CCTTTGTTCA CATTGCTTCT TCTTGGTGGA AGCTCTTCCT CTGGCCTTGA AAATGCCTGC 6240
        TTCTCTTCA AGGTAGCACA GTCATCACTT TCTGTGGTAA CCTTCTCCAG CACCATCAAA 6300
        CAGAAAGAAT GAATCTCTTG TAAATTCAGC TCTTACGTCA TTCATTACAT TATTTTGTAA 6360
        CTCTTTATAG ATTCTTCTCT CCCACTAGAC TCTGAGTCAC TGGAGAGTAG GAGCCAACTC 6420
        TCATTCATGT GTGGTTTGGT CAGCTACTGG CCACATTCCT GATGCATAGT TAATGCTCAA 6480 ACCTTAACTG GTGAATCAGC TCAAATATTG TCCTTCTCTA AATCCATTCA CTCATTGACT 6540 AACTATGTAC TCAAAATAGT AAACACCAGT AATTTAATCC AATTCCTGCC CATACTGCTT 6600 GGTACATTCC AGGTGAATTA GTTTGATAAA TATGTGTGTA TTACATAATA TTAAAGTATG 6660
15
        TACAGAAGAT CATGCTAATC ATAATTCACA ACTGATAACT AATCAAACAT AAATGCTCTC 6720
        AGGTTAACAA ATGTCTGCCT TCTCAGTTAA TGCAGTCATT AACAAACACC TTCTGATGCT 6780
        GATAATAGGG CCTTGTTCAG CAATGAAGCC ATAAAGGTGA ATAAAGAACA TGCCCTCGTG 6840
20
       GAGCTCACAG CCTAGTCATT ATTGTTCTGA TTTTTAATAT TAATGTTGGT TTGGGTTTTG 6900
GTGAAAAATG TTTAGACTTA TCTTAGTGAT CTTTTCATCC TTTGCTATAT TATTTTTCTC 6960
TAAGAGTCTT CCTTATCCCC TCCTTTAAAA AACTAGGTGA TAATTCTAAA TTGTAAATTT 7020
        AAATATTATA AATAGCTTAT AAAATTTAAT ATTTATAATA TTTAAATGTT TGATAAATAT 7080
        25
        TTTAAATGTG TTGGCCAGGC ATGGTGGCTG ACACCTATAA TCCCAGAACT TTGAGAGGCC 7200
       GGATGTTGCA GTGAGCTGAG ATCGTGCCAC TGCACTCCAA CCTGGGTGAC AGAGTGAGAC 7440
        TCCATCTCAA AAAAAAAAA TGTTATCTAA ATAAGATAAA TTTAATAACT GTTCGCACTT 7500
        AGATGAGCAT AAGGAACTAA ACCTAGATAA AACTATCAAA TAAGGCCTGG GTACAGTGAC 7560
        TCATGCCTGT AATCTCAAGC ACTTTGGGAG GCCAAAATTA TACAAAGTTA GTTGTATAAC 7620
       ACCAACTAAC AACTATTTTG GGGTTAGCTT AATTCAGATT AATTTTTTT AAACTGAGTT 7680
TTAAATTCCT GCTTACTCTA CCATACATGC TAGGCCTCAT ATTATGCTAG AAAAATTTTG 7740
AGCACAGATT TATGAATACT CTCCTGCATA CCATTTAATT TTTAAACAAA TTTTAATGCA 7800
       GTATATATGT GCCTTTTTAC CAACACATTA AATAATAAGA TCTACTGTGA GGACTAAATT 7860
35
       TCTGTAATTT CAAAGTAGTA ATGAGTTTAA ACCATGTCTC AAGATCTCTG CAATAACTGT 7920
       AGCACAACAG AAAATAGGTA TTTCTATTAA TGACAGAGTC ACAAGTACTA CTAATAATAC 7980 TGTGGTTTGT TTCCTGCAAC TAATCATGGG AGGAATGCTA AATTTCAGAG GTTGGTGAAA 8040 ATACATGTGT ATTTTTTCC CCATCCAAGT TCACAGATTT CTCACACTGA GAACTCCTAT 8100 TCCATAACAA AATTCTGGAA GCCTGCACAC CGTATTGGAA GAAGGGCAGA AAGGAAAAGC 8160
       AAATGGAAGG ATTTAAATTT TTTTCAAATC CTGTATCCCT TGATTTTACA GCAAGATTGT 8220
40
       ATTTATGTAT TACTTGTGTT AAAAATATAG TATAATCGAG ACTCCAGATC AAAAATCACC 8280
       GCAGCTCAGG GAGAAAGAGG GCCACCAAAT GCCAGAGCCC TTCAGCCTTC TCCCACCCTG 8340
       CCTGTACCCT CAGATGGAAG CACTTTTTTA TCATTGTTTC ACCTTTAGCA TTTTGACAAT 8400
GAAGTCACAA ACCTTCAGCC TCTCACCCAT AGGAACCCAC TGGTTGTAAG AGAAGGATGA 8460
AGCCAGTCCT TCCTAAAGGG CACGATTAGA TGTGTTTATG GCATCCTCAG GTGAAACTAT 8520
       45
       TACTAGGCCA TTTATCTACC CTTTATAATA TTGTTTAATG AGAAAATGCT TTCTATCTTC 8640
       CAAATATCTG ATGATTTGTA AGAGAACACT TAAACATGGG TATTCATAAG CTGAAACTTC 8700
       TGGCATTTAT TGAATGTCAA GATTGTTCAT CAGTATACTA GGTGATTAAC TGACCACTGA 8760
       ACTTGAAGGT AGTATAAAGT AGTAGTAAAA GGTACAATCA TTGTCTCTTA ACAGATGGCT 8820
       CTTTGCTTTC ATTAG
```

55

```
(B) TYPE: nucleic acid
                       (C) STRANDEDNESS: double
                      (D) TOPOLOGY: linear
                (ii) MOLECULE TYPE: Genomic DNA
 5
                (vi) ORIGINAL SOURCE:
                      (A) ORGANISM: human
                      (F) TISSUE TYPE:placenta
                (iX) FEATURE:
                      (A) NAME/KEY: intron
                      (B) LOCATION: 1..1371
 10
                      (C) IDENTIFICATION METHODS: E
                (xi) SEQUENCE DESCRIPTION: SEQ ID NO:11:
         GTAAGGCTAA TGCCATAGAA CAAATACCAG GTTCAGATAA ATCTATTCAA TTAGAAAAGA
         TGTTGTGAGG TGAACTATTA AGTGACTCTT TGTGTCACCA AATTTCACTG TAATATTAAT
         GGCTCTTAAA AAAATAGTGG ACCTCTAGAA ATTAACCACA ACATGTCCAA GGTCTCAGCA
15
                                                                                               180
        CCTTGTCACA CCACGTGTCC TGGCACTTTA ATCAGCAGTA GCTCACTCTC CAGTTGGCAG
TAAGTGCACA TCATGAAAAT CCCAGTTTTC ATGGGAAAAT CCCAGTTTTC ATTGGATTTC
CATGGGAAAA ATCCCAGTAC AAAACTGGGT GCATTCAGGA AATACAATTT CCCAAAGCAA
                                                                                               240
        ATTGGCAAAT TATGTAAGAG ATTCTCTAAA TTTAGAGTTC CGTGAATTAC ACCATTTTAT
GTAAATATGT TTGACAAGTA AAAATTGATT CTTTTTTTT TTTTCTGTTG CCCAGGCTGG
20
         AGTGCAGTGG CACAATCTCT GCTCACTGCA ACCTCCACCT CCTGGGTTCA AGCAATTCTC 540
         CTGCCTCAGC CTTCTGAGTA GCTGGGACTA CAGGTGCATC CCGCCATGCC TGGCTAATTT
                                                                                               600
         TTGGGTATTT TTACTAGAGA CAGGGTTTTG GCATGTTGTC CAGGCTGGTC TTGGACTCCT
        GATCTCAGAT GATCCTCCTG GCTCGGGCTC CCAAAGTGCT GGGATTACAG GCATGAACCA CCACACATGG CCTAAAAATT GATTCTTATG ATTAATCTCC TGTGAACAAT TTGGCTTCAT TTGAAAGTTT GCCTTCATTT GAAACCTTCA TTTAAAAGCC TGAGCAACAA AGTGAGACCC
        CATCTCTACA AAAAACTGCA AAATATCCTG TGGACACCTC CTACCTTCTG TGGAGGCTGA
                                                                                               900
        AGCAGGAGGA TCACTTGAGC CTAGGAATTT GAGCCTGCAG TGAGCTATGA TCCCACCCCT
                                                                                               960
        ACACTCCAGC CTGCATGACA GTAGACCCTG ACACACACA ACAAAAAAA ACCTTCATAA 1020
AAAATTATTA GTTGACTTTT CTTAGGTGAC TTTCCGTTTA AGCAATAAAT TTAAAAGTAA 1080
AATCTCTAAT TTTAGAAAAT TTATTTTTAG TTACATATTG AAATTTTTAA ACCCTAGGTT 1140
        TAAGTTTTAT GTCTAAATTA CCTGAGAACA CACTAAGTCT GATAAGCTTC ATTTTATGGG 1200
30
        CCTTTTGGAT GATTATATAA TATTCTGATG AAAGCCAAGA CAGACCCTTA AACCATAAAA 1260
        ATAGGAGTTC GAGAAAGAGG AGTAGCAAAA GTAAAAGCTA GAATGAGATT GAATTCTGAG 1320
        TCGAAATACA AAATTTTACA TATTCTGTTT CTCTCTTTTT CCCCCTCTTA G
        (13) INFORMATION FOR SEQ ID NO:12:
35
               (i) SEQUENCE CHARACTERISTICS:
                     (A) LENGTH: 3383 base pairs
                     (B) TYPE: nucleic acid
                     (C) STRANDEDNESS: double
                     (D) TOPOLOGY: linear
              (ii) MOLECULE TYPE: Genomic DNA
40
              (vi)ORIGINAL SOURCE:
                     (A) ORGANISM: human
                     (F) TISSUE TYPE:placenta
              (iX) FEATURE:
                     (A) NAME/KEY: intron
45
                     (B) LOCATION:1..3383
                     (C) IDENTIFICATION METHODS: E
              (xi) SEQUENCE DESCRIPTION: SEQ ID NO:12:
        GTAAAGTAGA AATGAATTTA TTTTTCTTTG CAAACTAAGT ATCTGCTTGA GACACATCTA
        TCTCACCATT GTCAGCTGAG GAAAAAAAA AATGGTTCTC ATGCTACCAA TCTGCCTTCA
50
       AAGAAATGTG GACTCAGTAG CACAGCTTTG GAATGAAGAT GATCATAAGA GATACAAAGA AGAACCTCTA GCAAAAGATG CTTCTCTATG CCTTAAAAAA TTCTCCAGCT CTTAGAATCT
                                                                                              240
       ACAAAATAGA CTTTGCCTGT TTCATTGGTC CTAAGATTAG CATGAAGCCA TGGATTCTGT
                                                                                              300
       TGTAGGGGGA GCGTTGCATA GGAAAAAGGG ATTGAAGCAT TAGAATTGTC CAAAATCAGT
                                                                                              360
       AACACCTCCT CTCAGAAATG CTTTGGGAAG AAGCCTGGAA GGTTCCGGGT TGGTGGTGGG
                                                                                             420
55
```

```
GTGGGGCAGA AAATTCTGGA AGTAGAGGAG ATAGGAATGG GTGGGGCAAG AAGACCACAT
          TCAGAGGCCA AAAGCTGAAA GAAACCATGG CATTTATGAT GAATTCAGGG TAATTCAGAA
         TGGAAGTAGA GTAGGAGTAG GAGACTGGTG AGAGGAGCTA GAGTGATAAA CAGGGTGTAG
          AGCAAGACGT TCTCTCACCC CAAGATGTGA AATTTGGACT TTATCTTGGA GATAATAGGG
5
         TTAATTAAGC ACAATATGTA TTAGCTAGGG TAAAGATTAG TTTGTTGTAA CAAAGACATC
         CAAAGATACA GTAGCTGAAT AAGATAGAGA ATTTTTCTCT CAAAGAAAGT CTAAGTAGGC
         AGCTCAGAAG TAGTATGGCT GGAAGCAACC TGATGATATT GGGACCCCCA ACCTTCTTCA
                                                                                                840
         GTCTTGTACC CATCATCCCC TAGTTGTTGA TCTCACTCAC ATAGTTGAAA ATCATCATAC
         TTCCTGGGTT CATATCCAG TTATCAAGAA AGGGTCAAGA GAAGTCAGGC TCATTCCTTT 960
CAAAGACTCT AATTGGAAGT TAAACACATC AATCCCCCTC ATATTCCATT GACTAGAATT 1020
TAATCACATG GCCACACCAA GTGCAAGGAA ATCTGGAAAA TATAATCTTT ATTCCAGGTA 1080
GCCATATGAC TCTTTAAAAT TCAGAAATAA TATATTTTTA AAATATCATT CTGGCTTTGG 1140
10
         TATAAAGAAT TGATGGTGTG GGGTGAGGAG GCCAAAATTA AGGGTTGAGA GCCTATTATT 1200
         TTAGTTATTA CAAGAAATGA TGGTGTCATG AATTAAGGTA GACATAGGGG AGTGCTGATG 1260
         AGGAGCTGTG AATGGATTTT AGAAACACTT GAGAGAATCA ATAGGACATG ATTTAGGGTT 1320
         GGATTTGGAA AGGAGAAGAA AGTAGAAAAG ATGATGCCTA CATTTTTCAC TTAGGCAATT 1380
15
         TGTACCATTC AGTGAAATAG GGAACACAGG AGGAAGAGCA GGTTTTGGTG TATACAAAGA 1440
GGAGGATGGA TGACGCATTT CGTTTTGGAT CTGAGATGTC TGTGGAACGT CCTAGTGGAG 1500
         ATGTCCACAA ACTCTTCTAC ATGTGGTTCT GAGTTCAGGA CACAGATTTG GGCTGGAGAT 1560
         AGAGATATTG TAGGCTTATA CATAGAAATG GCATTTGAAT CTATAGAGAT AAAAAGACAC 1620
         ATCAGAGGAA ATGTGTAAAG TGAGAGAGGA AAAGCCAAGT ACTGTGCTGG GGGGAATACC 1680
         TACATTTAAA GGATGCAGTA GAAAGAAGCT AATAAACAAC AGAGAGCAGA CTAACCAAAA 1740
GGGGAGAAGA AAAACCAAGA GAATTCCACC GACTCCCAGG AGAGCATTTC AAGATTGAGG 1800
20
         GGATAGGTGT TGTGTTGAAT TTTGCAGCCT TGAGAATCAA GGGCCAGAAC ACAGCTTTTA 1860
         GATTTAGCAA CAAGGAGTTT GGTGATCTCA GTGAAAGCAG CTTGATGGTG AAATGGAGGC 1920
AGAGGCAGAT TGCAATGAGT GAAACAGTGA ATGGGAAGTG AAGAAATGAT ACAGATAATT 1980
         CTTGCTAAAA GCTTGGCTGT TAAAAGGAGG AGAGAAACAA GACTAGCTGC AAAGTGAGAT 2040
25
         TGGGTTGATG GAGCAGTTTT AAATCTCAAA ATAAAGAGCT TTGTGCTTTT TTGATTATGA 2100
         AAATAATGTG TTAATTGTAA CTAATTGAGG CAATGAAAAA AGATAATAAT ATGAAAGATA 2160
AAAATATAAA AACCACCAG AAATAATGAT AGCTACCATT TTGATACAAT ATTTCTACAC 2220
TCCTTTCTAT GTATATATAC AGACACAGAA ATGCTTATAT TTTTATTAAA AGGGATTGTA 2280
         CTATACCTAA GCTGCTTTTT CTAGTTAGTG ATATATATGG ACATCTCTCC ATGGCAACGA 2340
         GTAATTGCAG TTATATTAAG TTCATGATAT TTCACAATAA GGGCATATCT TTGCCCTTTT 2400
30
         TATTTAATCA ATTCTTAATT GGTGAATGTT TGTTTCCAGT TTGTTGTTGT TATTAACAAT 2460
         GTTCCCATAA GCATTCCTGT ACACCAATGT TCACACATTT GTCTGATTTT TTCTTCAGGA 2520 TAAAACCCAG GAGGTAGAAT TGCTGGGTTG ATAGAAGAGA AAGGATGATT GCCAAATTAA 2580
         AGCTTCAGTA GAGGGTACAT GCCGAGCACA AATGGGATCA GCCCTAGATA CCAGAAATGG 2640
         CACTTTCTCA TTTCCCCTTG GGACAAAAGG GAGAGAGGCA ATAACTGTGC TGCCAGAGTT 2700
         AAATTTGTAC GTGGAGTAGC AGGAAATCAT TTGCTGAAAA TGAAAACAGA GATGATGTTG 2760
35
         TAGAGGTCCT GAAGAGAGCA AAGAAAATTT GAAATTGCGG CTATCAGCTA TGGAAGAGAG 2820
         TGCTGAACTG GAAAACAAAA GAAGTATTGA CAATTGGTAT GCTTGTAATG GCACCGATTT 2880
        GAACGCTTGT GCCATTGTTC ACCAGCAGCA CTCAGCAGCC AAGTTTGGAG TTTTGTAGCA 2940
GAAAGACAAA TAAGTTAGGG ATTTAATATC CTGGCCAAAT GGTAGACAAA ATGAACTCTG 3000
         AGATCCAGCT GCACAGGGAA GGAAGGGAAG ACGGGAAGAG GTTAGATAGG AAATACAAGA 3060
         GTCAGGAGAC TGGAAGATGT TGTGATATTT AAGAACACAT AGAGTTGGAG TAAAAGTGTA 3120
40
         AGAAAACTAG AAGGGTAAGA GACCGGTCAG AAAGTAGGCT ATTTGAAGTT AACACTTCAG 3180
        AGGCAGAGTA GTTCTGAATG GTAACAAGAA ATTGAGTGTG CCTTTGAGAG TAGGTTAAAA 3240
        AACAATAGGC AACTTTATTG TAGCTACTTC TGGAACAGAA GATTGTCATT AATAGTTTTA 3300
GAAAACTAAA ATATATAGCA TACTTATTTG TCAATTAACA AAGAAACTAT GTATTTTTAA 3360
        ATGAGATTTA ATGTTTATTG TAG
                                                                                              3383
```

(14) INFORMATION FOR SEQ ID NO:13:
 (i)SEQUENCE CHARACTERISTICS:
 (A)LENGTH:11464 base pairs
 (B)TYPE:nucleic acid

(C) STRANDEDNESS: double (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: Genomic DNA

(vi)ORIGINAL SOURCE:
 (A)ORGANISM:human

55

45

```
(F) TISSUE TYPE: placenta
             (iX) FEATURE:
                  (A) NAME/KEY:5'UTR
                  (B) LOCATION:1..3
5
                  (C) IDENTIFICATION METHODS: E
                  (A) NAME/KEY:leader peptide
                  (B) LOCATION: 4..82
                  (C) IDENTIFICATION METHODS:S
                  (A) NAME/KEY: intron
                  (B) LOCATION: 83..1453
10
                  (C) IDENTIFICATION METHODS: E
                  (A) NAME/KEY:leader peptide
                  (B) LOCATION: 1454..1465
                  (C) IDENTIFICATION METHODS:S
                  (A) NAME/KEY: intron
                  (B) LOCATION: 1466..4848
15
                  (C) IDENTIFICATION METHODS: E
                  (A) NAME/KEY:leader peptide
                  (B) LOCATION: 4849..4865
                  (C) IDENTIFICATION METHODS:S
                  (A) NAME/KEY: mat peptide
20
                  (B) LOCATION: 4866..4983
                  (C) IDENTIFICATION METHODS:S
                  (A) NAME/KEY: intron
                  (B) LOCATION: 4984..6317
                  (C) IDENTIFICATION METHODS: E
                  (A) NAME/KEY: mat peptide
25
                  (B) LOCATION: 6318..6451
                  (C) IDENTIFICATION METHODS:S
                  (A) NAME/KEY: intron
                  (B) LOCATION: 6452..11224
                 (C) IDENTIFICATION METHODS: E
                  (A) NAME/KEY:mat peptide
30
                  (B) LOCATION: 11225..11443
                  (C) IDENTIFICATION METHODS:S
                  (A) NAME/KEY: 3'UTR
                 (B) LOCATION: 11444..11464
                 (C) IDENTIFICATION METHODS: E
            (xi) SEQUENCE DESCRIPTION: SEQ ID NO:13:
35
      AAG ATG GCT GCT GAA CCA GTA GAA GAC AAT TGC ATC AAC TTT GTG GCA
          Met Ala Ala Glu Pro Val Glu Asp Asn Cys Ile Asn Phe Val Ala
               -35
                                    -30
      ATG AAA TTT ATT GAC AAT ACG CTT TAC TTT ATA G
                                                          GTAAGG CTAATGCCAT
40
      98
      Met Lys Phe Ile Asp Asn Thr Leu Tyr Phe Ile Ala
           -20
                                -15
                                                      -10
      AGAACAAATA CCAGGTTCAG ATAAATCTAT TCAATTAGAA AAGATGTTGT GAGGTGAACT
      ATTAAGTGAC TCTTTGTGTC ACCAAATTTC ACTGTAATAT TAATGGCTCT TAAAAAAATA
45
      GTGGACCTCT AGAAATTAAC CACAACATGT CCAAGGTCTC AGCACCTTGT CACACCACGT
      GTCCTGGCAC TTTAATCAGC AGTAGCTCAC TCTCCAGTTG GCAGTAAGTG CACATCATGA
50
      AAATCCCAGT TTTCATGGGA AAATCCCAGT TTTCATTGGA TTTCCATGGG AAAAATCCCA
      GTACAAAACT GGGTGCATTC AGGAAATACA ATTTCCCAAA GCAAATTGGC AAATTATGTA
      AGAGATTCTC TAAATTTAGA GTTCCGTGAA TTACACCATT TTATGTAAAT ATGTTTGACA
```

24

	518 AGTAAAAATT	GATTCTTTT	TTTTTTTTCT	· GTTGCCCAGG	CTGGAGTGCA GTGGCACAAT	
	578				TCTCCTGCCT CAGCCTTCTG	
5	638					
	698				ATTTTTGGGT ATTTTTACTA	
	GAGACAGGGT 758	TTTGGCATGT	TGTCCAGGCT	GGTCTTGGAC	TCCTGATCTC AGATGATCCT	
10	CCTGGCTCGG 818	GCTCCCAAAG	TGCTGGGATT	ACAGGCATGA	ACCACCACAC ATGGCCTAAA	
	AATTGATTCT 878	' TATGATTAAT	CTCCTGTGAA	CAATTTGGCT	TCATTTGAAA GTTTGCCTTC	
	ATTTGAAACC	TTCATTTAAA	AGCCTGAGCA	ACAAAGTGAG	ACCCCATCTC TACAAAAAC	
15	998				CTGAAGCAGG AGGATCACTT	
	1058			•	CCCTACACTC CAGCCTGCAT	
	GACAGTAGAC 1118	CCTGACACAC	ACACACAAAA	AAAAACCTTC	ATAAAAAATT ATTAGTTGAC	
20	1178				GTAAAATCTC TAATTTTAGA	
	1238				GGTTTAAGTT TTATGTCTAA	
	ATTACCTGAG 1298	AACACACTAA	GTCTGATAAG	CTTCATTTTA	TGGGCCTTTT GGATGATTAT	
25	ATAATATTCT 1358	GATGAAAGCC	AAGACAGACC	CTTAAACCAT	AAAAATAGGA GTTCGAGAAA	
	GAGGAGTAGC 1418	AAAAGTAAAA	GCTAGAATGA	GATTGAATTC	TGAGTCGAAA TACAAAATTT	
		GTTTCTCTCT	TTTTCCCCCT	CTTAG CT	GAA GAT GAT G GTAAA	
30					Glu Asp Asp Glu	
	GTAGAAATGA 1530	ATTTATTTT	CTTTGCAAAC	TAAGTATCTG	CTTGAGACAC ATCTATCTCA	
		CTGAGGAAAA	AAAAAAATGG	TTCTCATGCT	ACCAATCTGC CTTCAAAGAA	
35	1590 ATGTGGACTC				ACCAATCTGC CTTCAAAGAA TAAGAGATAC AAAGAAGAAC	
35	1590 ATGTGGACTC 1650 CTCTAGCAAA	AGTAGCACAG	CTTTGGAATG	AAGATGATCA		
	1590 ATGTGGACTC 1650 CTCTAGCAAA 1710 ATAGACTTTG	AGTAGCACAG AGATGCTTCT	CTTTGGAATG CTATGCCTTA	AAGATGATCA AAAAATTCTC	TAAGAGATAC AAAGAAGAAC	
<i>35</i>	1590 ATGTGGACTC 1650 CTCTAGCAAA 1710 ATAGACTTTG 1770 GGGGAGCGTT	AGTAGCACAG AGATGCTTCT CCTGTTTCAT	CTTTGGAATG CTATGCCTTA TGGTCCTAAG	AAGATGATCA AAAAATTCTC ATTAGCATGA	TAAGAGATAC AAAGAAGAAC CAGCTCTTAG AATCTACAAA	
	1590 ATGTGGACTC 1650 CTCTAGCAAA 1710 ATAGACTTTG 1770 GGGGAGCGTT 1830 CTCCTCTCAG	AGTAGCACAG AGATGCTTCT CCTGTTTCAT GCATAGGAAA	CTTTGGAATG CTATGCCTTA TGGTCCTAAG AAGGGATTGA	AAGATGATCA AAAAATTCTC ATTAGCATGA AGCATTAGAA	TAAGAGATAC AAAGAAGAAC CAGCTCTTAG AATCTACAAA AGCCATGGAT TCTGTTGTAG	
40	1590 ATGTGGACTC 1650 CTCTAGCAAA 1710 ATAGACTTTG 1770 GGGGAGCGTT 1830 CTCCTCTCAG 1890 GCAGAAAATT	AGTAGCACAG AGATGCTTCT CCTGTTTCAT GCATAGGAAA AAATGCTTTG	CTTTGGAATG CTATGCCTTA TGGTCCTAAG AAGGGATTGA GGAAGAAGCC	AAGATGATCA AAAAATTCTC ATTAGCATGA AGCATTAGAA TGGAAGGTTC	TAAGAGATAC AAAGAAGAAC CAGCTCTTAG AATCTACAAA AGCCATGGAT TCTGTTGTAG TTGTCCAAAA TCAGTAACAC	
	1590 ATGTGGACTC 1650 CTCTAGCAAA 1710 ATAGACTTTG 1770 GGGGAGCGTT 1830 CTCCTCTCAG 1890 GCAGAAAATT 1950 GGCCAAAAGC	AGTAGCACAG AGATGCTTCT CCTGTTTCAT GCATAGGAAA AAATGCTTTG CTGGAAGTAG	CTTTGGAATG CTATGCCTTA TGGTCCTAAG AAGGGATTGA GGAAGAAGCC AGGAGATAGG	AAGATGATCA AAAAATTCTC ATTAGCATGA AGCATTAGAA TGGAAGGTTC AATGGGTGGG	TAAGAGATAC AAAGAAGAAC CAGCTCTTAG AATCTACAAA AGCCATGGAT TCTGTTGTAG TTGTCCAAAA TCAGTAACAC CGGGTTGGTG GTGGGGTGGG	
40	1590 ATGTGGACTC 1650 CTCTAGCAAA 1710 ATAGACTTTG 1770 GGGGAGCGTT 1830 CTCCTCTCAG 1890 GCAGAAAATT 1950 GGCCAAAAGC 2010	AGTAGCACAG AGATGCTTCT CCTGTTTCAT GCATAGGAAA AAATGCTTTG CTGGAAGTAG TGAAAGAAAC	CTTTGGAATG CTATGCCTTA TGGTCCTAAG AAGGGATTGA GGAAGAAGCC AGGAGATAGG CATGGCATTT	AAGATGATCA AAAAATTCTC ATTAGCATGA AGCATTAGAA TGGAAGGTTC AATGGGTGGG ATGATGAATT	TAAGAGATAC AAAGAAGAAC CAGCTCTTAG AATCTACAAA AGCCATGGAT TCTGTTGTAG TTGTCCAAAA TCAGTAACAC CGGGTTGGTG GTGGGGTGGG GCAAGAAGAC CACATTCAGA	
40	1590 ATGTGGACTC 1650 CTCTAGCAAA 1710 ATAGACTTTG 1770 GGGGAGCGTT 1830 CTCCTCTCAG 1890 GCAGAAAATT 1950 GGCCAAAAGC 2010 GTAGAGTAGG 2070 GACGTTCTCT	AGTAGCACAG AGATGCTTCT CCTGTTTCAT GCATAGGAAA AAATGCTTTG CTGGAAGTAG TGAAAGAAAC AGTAGGAGAC	CTTTGGAATG CTATGCCTTA TGGTCCTAAG AAGGGATTGA GGAAGAAGCC AGGAGATAGG CATGGCATTT TGGTGAGAGG	AAGATGATCA AAAAATTCTC ATTAGCATGA AGCATTAGAA TGGAAGGTTC AATGGGTGGG ATGATGAATT AGCTAGAGTG	TAAGAGATAC AAAGAAGAAC CAGCTCTTAG AATCTACAAA AGCCATGGAT TCTGTTGTAG TTGTCCAAAA TCAGTAACAC CGGGTTGGTG GTGGGGTGGG GCAAGAAGAC CACATTCAGA CAGGGTAATT CAGAATGGAA	
40	1590 ATGTGGACTC 1650 CTCTAGCAAA 1710 ATAGACTTTG 1770 GGGGAGCGTT 1830 CTCCTCTCAG 1890 GCAGAAAATT 1950 GGCCAAAAGC 2010 GTAGAGTAGG 2070 GACGTTCTCT 2130	AGTAGCACAG AGATGCTTCT CCTGTTTCAT GCATAGGAAA AAATGCTTTG CTGGAAGTAG TGAAAGAAAC AGTAGGAGAC CACCCCAAGA	CTTTGGAATG CTATGCCTTA TGGTCCTAAG AAGGGATTGA GGAAGAAGCC AGGAGATAGG CATGGCATTT TGGTGAGAGG TGTGAAATTT	AAGATGATCA AAAAATTCTC ATTAGCATGA AGCATTAGAA TGGAAGGTTC AATGGGTGGG ATGATGAATT AGCTAGAGTG GGACTTTATC	TAAGAGATAC AAAGAAGAAC CAGCTCTTAG AATCTACAAA AGCCATGGAT TCTGTTGTAG TTGTCCAAAA TCAGTAACAC CGGGTTGGTG GTGGGGTGGG GCAAGAAGAC CACATTCAGA CAGGGTAATT CAGAATGGAA ATAAACAGGG TGTAGAGCAA	
40	1590 ATGTGGACTC 1650 CTCTAGCAAA 1710 ATAGACTTTG 1770 GGGGAGCGTT 1830 CTCCTCTCAG 1890 GCAGAAAATT 1950 GGCCAAAAGC 2010 GTAGAGTAGG 2070 GACGTTCTCT 2130 TAAGCACAAT 2190	AGTAGCACAG AGATGCTTCT CCTGTTTCAT GCATAGGAAA AAATGCTTTG CTGGAAGTAG TGAAAGAAAC AGTAGGAGAC CACCCCAAGA ATGTATTAGC	CTTTGGAATG CTATGCCTTA TGGTCCTAAG AAGGGATTGA GGAAGAAGCC AGGAGATAGG CATGGCATTT TGGTGAGAGG TGTGAAATTT TAGGGTAAAG	AAGATGATCA AAAAATTCTC ATTAGCATGA AGCATTAGAA TGGAAGGTTC AATGGGTGGG ATGATGAATT AGCTAGAGTG GGACTTTATC ATTAGTTTGT	TAAGAGATAC AAAGAAGAAC CAGCTCTTAG AATCTACAAA AGCCATGGAT TCTGTTGTAG TTGTCCAAAA TCAGTAACAC CGGGTTGGTG GTGGGGTGGG GCAAGAAGAC CACATTCAGA CAGGGTAATT CAGAATGGAA ATAAACAGGG TGTAGAGCAA TTGGAGATAA TAGGGTTAAT	

	AGAAGTAGTA 2310	TGGCTGGAAG	CAACCTGATG	ATATTGGGAC	CCCCAACCTT	CTTCAGTCTT
		TCCCCTAGTT	GTTGATCTCA	CTCACATAGT	TGAAAATCAT	CATACTTCCT
5		CCCAGTTATC	AAGAAAGGGT	CAAGAGAAGT	CAGGCTCATT	CCTTTCAAAG
		GAAGTTAAAC	ACATCAATCC	CCCTCATATT	CCATTGACTA	GAATTTAATC
10		ACCAAGTGCA	AGGAAATCTG	GAAAATATAA	TCTTTATTCC	AGGTAGCCAT
	ATGACTCTTT 2610	AAAATTCAGA	AATAATATAT	TTTTAAAATA	TCATTCTGGC	TTTGGTATAA
	AGAATTGATG 2670	GTGTGGGGTG	AGGAGGCCAA	AATTAAGGGT	TGAGAGCCTA	TTATTTTAGT
15	TATTACAAGA 2730	AATGATGGTG	TCATGAATTA	AGGTAGACAT	AGGGGAGTGC	TGATGAGGAG
	CTGTGAATGG 2790	ATTTTAGAAA	CACTTGAGAG	AATCAATAGG	ACATGATTTA	GGGTTGGATT
	TGGAAAGGAG 2850	AAGAAAGTAG	AAAAGATGAT	GCCTACATTT	TTCACTTAGG	CAATTTGTAC
20	2910	AATAGGGAAC				
	2970	CATTTCGTTT				
	3030	TCTACATGTG				
25	3090	TTATACATAG				
	3150	TAAAGTGAGA				
	3210	CAGTAGAAAG				
30	3270	CAAGAGAATT				
	3330	TGAATTTTGC				
05	3390	AGTTTGGTGA				
35	3450	TGAGTGAAAC				
	3510	GCTGTTAAAA				
40	3570	GTTTTAAATC				
	3630	TGTAACTAAT				
	3690	CCCAGAAATA				
45	3750	TATACAGACA				
	3810	TTTTTCTAGT				
	3870	TTAAGTTCAT				
50	3930	TAATTGGTGA				
	3990	CCTGTACACC AGAATTGCTG				
	4050	TACATGCCGA				
55	CHOINGHOOG	INCMIGCEGA	GCACAAAIGG	GAICAGUCUT	AGATACCAGA	AATGGCACTT

	4110					
		CTTGGGACA	AAAGGGAGAG	AGGCAATAAC	TGTGCTGCCA G	AGTTAAATT
	4170	-1.CC1.CC1.1	<b>1 201 200</b>			
5	4230	TAGCAGGAA	ATCATTTGCT	GAAAATGAAA	ACAGAGATGA T	GTTGTAGAG
	GTCCTGAAGA G	Ασσααασρα	<b>ልልጥጥጥ</b> ሮል ልልጥ	тессестате	ACCEATCOA A	ACACMCOMO
	4290		AATTIOAAA:	IGCGGCIAIC	AGCIAIGGAA G	AGAGIGCIG
	AACTGGAAAA C	AAAAGAAGT	ATTGACAATT	GGTATGCTTG	TAATGGCACC G	ATTTGAACG
	4350					
10	CTTGTGCCAT TO	GTTCACCAG	CAGCACTCAG	CAGCCAAGTT	TGGAGTTTTG T	'AGCAGAAAG
	4410					
	ACAAATAAGT TA	AGGGATTTA .	ATATCCTGGC	CAAATGGTAG	ACAAAATGAA C	TCTGAGATC
	CAGCTGCACA GO	GAAGGAAG	GGAAGACGGG	AACACCTTAC	ATAGGAAATA C	3 3 C 3 C T C 3 C
	4530		OURNORCOGO	AAGAGGTTAG	AIAGGAAIA C	AAGAGICAG
15	GAGACTGGAA GA	ATGTTGTGA	TATTTAAGAA	CACATAGAGT	TGGAGTAAAA G	TGTAAGAAA
	4590					
	ACTAGAAGGG TA	AAGAGACCG (	GTCAGAAAGT	AGGCTATTTG	AAGTTAACAC T	TCAGAGGCA
	4650	`				
	GAGTAGTTCT GA	AAIGGTAAC A	AAGAAATTGA	GTGTGCCTTT	GAGAGTAGGT T.	AAAAAACAA
20	TAGGCAACTT TA	ATTGTAGCT	ΑΓΤΤΟΤΙΙΑ	CAGAAGATTG	ጥሮልጥጥልልጥልር ጥ	<b>ጥጥጥስር</b> አአአአ
	4770			chormonito	ICATIANIAG I	IIIAGAAAA
	CTAAAATATA TA	GCATACTT A	ATTTGTCAAT	TAACAAAGAA	ACTATGTATT T	TTAAATGAG
	4830					
	ATTTAATGTT TA	ATTGTAG A	A AAC CTG (	GAA TCA GAT	TAC TTT GGC	AAG CTT
25	4880	C),	a Aco tou (	Tin Com Non	Tree Dho Clar	
		GI	-5	siu ser Asp	Tyr Phe Gly 1	Lys Leu 5
	GAA TCT AAA T	TA TCA GTO		AT TTG AAT	GAC CAA GTT	CTC TTC
	4928					
	Glu Ser Lys L		l Ile Arg A		Asp Gln Val 1	Leu Phe
30	ATT GAC CAA G	10 'Ca aan ccc	~ ~~~ ~~~ ~~~ ~~	15		20
	4976	GA AAI CGC	J CCI CIA I	II GAA GAI	AIG ACT GAT	ICI GAC
	Ile Asp Gln G	lv Asn Arc	Pro Leu P	he Glu Asp	Met Thr Asp S	Ser Asp
		5		0	35	p
		TATTTTTTT 1	TAATTCGCA	AACATAGAAA	TGACTAGCTA CT	TTCTTCCCA
<i>35</i>	5032					
	Cys Arg Asp 40					
	TTCTGTTTTA CT	GCTTACAT T	TGTTCCGTGC	TAGTCCCAAT	CCTCAGATGA AZ	AAGTCACAG
	5092			11101000111		mio i cheno
	GAGTGACAAT AA	TTTCACTT A	CAGGAAACT	TTATAAGGCA	TCCACGTTTT TT	PAGTTGGGG
40	5152					
	TAAAAAATTG GA	TACAATAA G	BACATTGCTA	GGGGTCATGC	CTCTCTGAGC CI	rgcctttga
	5212 ATCACCAATC CC	ייים אייים אייים מייים	፣ አጥጥ <i>ር</i> ር አጥጥ አ	አ ርጥር ጥጥጥ እ እ እ	ACCTCTATAC OF	nca a ma amm
	5272	TITALIGI G	MIIGCALIA	ACIGITIAAA	ACCICIAIAG II	IGGAIGCII
	AATCCCTGCT TG	TTACAGCT G	BAAAATGCTG	ATAGTTTACC	AGGTGTGGTG GO	CATCTATCT
45	5332					
	GTAATCCTAG CT	ACTTGGGA G	GCTCAAGCA	GGAGGATTGC	TTGAGGCCAG GA	ACTTTGAGG
	5392	CTC > TCC =	CCDCCC			
	CTGTAGTACA CT	GIGATUGT A	CCTGTGAAT	AGCCACTGCA	CTCCAGCCTG GG	STGATATAC
	AGACCTTGTC TC	ΤΑΑΑΑΤΤΑ Δ	ασασασα		СТТАССАААС С	ልልሞሞርልሞር
50	5512				C.IACOAAAG GA	EMITONIC
	AAGTCTACTG TG	CCTTCCAA A	ACATGAATT	CCAAATATCA	AAGTTAGGCT GA	GTTGAAGC
	5572	mm.cmm				
	AGTGAATGTG CA	TTCTTTAA A	AATACTGAA	TACTTACCTT	AACATATATT TI	TAAATATTT

```
TATTTAGCAT TTAAAAGTTA AAAACAATCT TTTAGAATTC ATATCTTTAA AATACTCAAA
     5692
     AAAGTTGCAG CGTGTGTTT GTAATACACA TTAAACTGTG GGGTTGTTTG TTTGTTTGAG
     5752
5
     5812
     CTCACTACAA CCTCCACCTC CCACGTTCAA GCGATTCTCA TGCCTCAGTC TCCCGAGTAG
     5872
    GTGGGATTAC AGGCATGCAC CACTTACACC CGGCTAATTT TTGTATTTTT AGTAGAGCTG
     5932
10
    GGGTTTCACC ATGTTGGCCA GGCTGGTCTC AAACCCCTAA CCTCAAGTGA TCTGCCTGCC
     5992
    TCAGCCTCCC AAACAAACAA ACAACCCCAC AGTTTAATAT GTGTTACAAC ACACATGCTG
    6052
    CAACTTTTAT GAGTATTTTA ATGATATAGA TTATAAAAGG TTGTTTTTAA CTTTTAAATG
15
     6112
    CTGGGATTAC AGGCATGAGC CACTGTGCCA GGCCTGAACT GTGTTTTTAA AAATGTCTGA
    6172
    CCAGCTGTAC ATAGTCTCCT GCAGACTGGC CAAGTCTCAA AGTGGGAACA GGTGTATTAA
    6232
    GGACTATCCT TTGGTTAAAT TTCCGCAAAT GTTCCTGTGC AAGAATTCTT CTAACTAGAG
20
    6292
    TTCTCATTTA TTATATTTAT TTCAG AT AAT GCA CCC CGG ACC ATA TTT ATT
    6343
                                Asp Asn Ala Pro Arg Thr Ile Phe Ile
                                40
                                                    45
    ATA AGT ATG TAT AAA GAT AGC CAG CCT AGA GGT ATG GCT GTA ACT ATC
25
    6391
    Ile Ser Met Tyr Lys Asp Ser Gln Pro Arg Gly Met Ala Val Thr Ile
        50
    TCT GTG AAG TGT GAG AAA ATT TCA ACT CTC TCC TGT GAG AAC AAA ATT
    6439
    Ser Val Lys Cys Glu Lys Ile Ser Thr Leu Ser Cys Glu Asn Lys Ile
30
    55
    ATT TCC TTT AAG GTAAG ACTGAGCCTT ACTTTGTTTT CAATCATGTT AATATAATCA
    6496
    Ile Ser Phe Lys
    ATATAATTAG AAATATAACA TTATTTCTAA TGTTAATATA AGTAATGTAA TTAGAAAACT
35
    6556
    CAAATATCCT CAGACCAACC TTTTGTCTAG AACAGAAATA ACAAGAAGCA GAGAACCATT
    6616
    AAAGTGAATA CTTACTAAAA ATTATCAAAC TCTTTACCTA TTGTGATAAT GATGGTTTTT
    6676
    CTGAGCCTGT CACAGGGGAA GAGGAGATAC AACACTTGTT TTATGACCTG CATCTCCTGA
40
    6736
    ACAATCAGTC TTTATACAAA TAATAATGTA GAATACATAT GTGAGTTATA CATTTAAGAA
    6796
    TAACATGTGA CTTTCCAGAA TGAGTTCTGC TATGAAGAAT GAAGCTAATT ATCCTTCTAT
    6856
    ATTTCTACAC CTTTGTAAAT TATGATAATA TTTTAATCCC TAGTTGTTTT GTTGCTGATC
45
    6916
    CTTAGCCTAA GTCTTAGACA CAAGCTTCAG CTTCCAGTTG ATGTATGTTA TTTTTAATGT
    6976
    TAATCTAATT GAATAAAAGT TATGAGATCA GCTGTAAAAG TAATGCTATA ATTATCTTCA
    7036
    AGCCAGGTAT AAAGTATTTC TGGCCTCTAC TTTTTCTCTA TTATTCTCCA TTATTATTCT
50
    7096
    CTATTATTT TCTCTATTTC CTCCATTATT GTTAGATAAA CCACAATTAA CTATAGCTAC
    7156
    AGACTGAGCC AGTAAGAGTA GCCAGGGATG CTTACAAATT GGCAATGCTT CAGAGGAGAA
    7216
```

	TTCCATGTCA	TGAAGACTCT	TTTTGAGTGG	AGATTTGCCA	ATAAATATCC	GCTTTCATGC
	CCACCCAGTC 7336	CCCACTGAAA	GACAGTTAGG	ATATGACCTT	AGTGAAGGTA	CCAAGGGGCA
5		GAGAAAAAAG	CCACTCTAAA	ATATAATCCA	AGTAAGAACA	GTGCATATGC
		GCCCCAGAC	AAATCCCTCA	GCTATCTCCC	TCCAACCAGA	GTGCCACCCC
10		AATTTGGAGT	CCCCATTCTA	GACCTGACAG	GCAGCTTAGT	TATCAAAATA
70		CCTGGGATGG	AAGGGTAGGG	TGGAAAGGGT	TAAGCATGCT	GTTACTGAAC
	AACATAATTA 7636	GAAGGGAAGG	AGATGGCCAA	GCTCAAGCTA	TGTGGGATAG	AGGAAAACTC
15		GCAGATTCAG	AAACTGGGAT	AAGTCCGAAC	CTACAGGTGG	ATTCTTGTTG
		GTGAAAATGT	TAAGAAGATG	GAAATAATGC	TTGGCACTTA	GTAGGAACTG
	GGCAAATCCA 7816	TATTTGGGGG	AGCCTGAAGT	TTATTCAATT	TTGATGGCCC	TTTTAAATAA
20	AAAGAATGTG 7876	GCTGGGCGTG	GTGGCTCACA	CCTGTAATCC	CAGCACTTTG	GGAGGCCGAG
	GGGGGCGGAT	CACCTGAAGT	CAGGAGTTCA	AGACCAGCCT	GACCAACATG	GAGAAACCCC
	ATCTCTACTA 7996	AAAATACAAA	ATTAGCTGGG	CGTGGTGGCA	TATGCCTGTA	ATCCCAGCTA
25	CTCGGGAGGC 8056	TGAGGCAGGA	GAATCTTTTG	AACCCGGGAG	GCAGAGGTTG	CGATGAGCCT
	AGATCGTGCC 8116	ATTGCACTCC	AGCCTGGGCA	ACAAGAGCAA	AACTCGGTCT	СААААААА
	AAAAAAAAAG 8176	TGAAATTAAC	CAAAGGCATT	AGCTTAATAA	TTTAATACTG	TTTTTAAGTA
· 30	GGGCGGGGGG 8236	TGGCTGGAAG	AGATCTGTGT	AAATGAGGGA	ATCTGACATT	TAAGCTTCAT
	CAGCATCATA 8296	GCAAATCTGC	TTCTGGAAGG	AACTCAATAA	ATATTAGTTG	GAGGGGGGA
	GAGAGTGAGG 8356	GGTGGACTAG	GACCAGTTTT	AGCCCTTGTC	TTTAATCCCT	TTTCCTGCCA
35	CTAATAAGGA 8416	TCTTAGCAGT	GGTTATAAAA	GTGGCCTAGG	TTCTAGATAA	TAAGATACAA
	CAGGCCAGGC 8476	ACAGTGGCTC	ATGCCTATAA	TCCCAGCACT	TTGGGAGGGC	AAGGCGAGTG
	8536	GATCAGGAGT				
40	CTAAAAAAAA 8596	TACAAAAATT	AGCCAGGCAT	GGTGGCATGC	ACCTGTAATC	CCAGCTACTC
	GTGAGCCTGA 8656	GGCAGAAGAA	TCGCTTGAAA	CCAGGAGGTG	TAGGCTGCAG	TGAGCTGAGA
45	8716					AAAAGAAAAA
40	8776	GCTACCCTTA			-	
	TATAAAGTTC 8836	TTTGGTCAAG	AACCTTGACA	ACACTAAGAG	GGATTTGCTT	TGAGAGGTTA
50	8896	CTGTTTCATA				
	8956					ATATCAACTT
	9016	CAGATTCAAC				
55	CCTTTCCCCT	TGGAGCACTC	AAGTTTCACC	AGGTGGGGCT	TTCCAAGTTG	GGGGTTCTCC

	9076					
	9136					ATGGCTGGGA
5	9196					CTATTCTGAA
	ATGTGCAATA 9256	AGTGTGATTA	AAGAGATTGC	CTGTTCTACC	TATCCACACT	CTCGCTTTCA
	ACTGTAACTT	TCTTTTTTC	TTTTTTTCTT	TTTTTCTTTT	TTTTTGAAAC	GGAGTCTCGC
10		AGGCTAGAGT	GCAGTGGCAC	GATCTCAGCT	CACTGCAAGC	TCTGCCTCCC
		CATTCTCCTG	CCTCACCCTC	CCAAGCAGCT	GGGACTACAG	GCGCCTGCCA
		CTAATTTTT	GTATTTTTAG	TAGAGACGGG	GTTTCACCGT	GTTAGCCAGG
15		TCTCCTGAAC	TTGTGATCCG	CCCGCCTCAG	CCTCCCAAAG	TGCTGGGATT
		GCCATCGCAC	CCGGCTCAAC	TGTAACTTTC	TATACTGGTT	CATCTTCCCC
		CTAGAGCTTT	TGAAGTTTTG	GCTATGGATT	ATTTCTCATT	TATACATTAG
20		AGTTCCAAAT	TGATGCCCAC	AGCTTAGGGT	CTCTTCCTAA	ATTGTATATT
		GCAGAAGTGG	GTGCCAATAG	GGGAACTAGT	TTATACTTTC	ATCAACTTAG
		TGTTGATAAA	GAACAAAGGT	CAAGAGTTAT	GACTACTGAT	TCCACAACTG
25		TGGAGATAAC	CCCGTGACCT	CTGCCATCCA	GAGTCTTTCA	GGCATCTTTG
		AAATGCTATT	TTAATTTTGG	AGGTTTCTCT	ATCAGTGCTT	AGGATCATGG
		TGCCATGAGG	CCAAAATTAA	GTCCAAAACA	TCTACTGGTT	CCAGGATTAA
30		CCTTAGGTGG	TGCCCACATG	TTCTGATCCA	TCCTGCAAAA	TAGACATGCT
		GAAAAGTGCA	GGCAGCACTA	CCAGTTGGAT	AACCTGCAAG	ATTATAGTTT
		AACCATTTCT	CACAAGGCCC	TATTCTGTGA	CTGAAACATA	CAAGAATCTG
35		TCTAAGGCAG	GGCCCAGCCA	AGGAGACCAT	ATTCAGGACA	GAAATTCAAG
	ACTACTATGG	AACTGGAGTG	CTTGGCAGGG	AAGACAGAGT	CAAGGACTGC	CAACTGAGCC
		GCTTACACAG	GAACCCAGGG	CCTAGCCCTA	CAACAATTAT	TGGGTCTATT
40		TTTAATTTCA	GGCTCCACTG	AAAGAGTAAG	CTAAGATTCC	TGGCACTTTC
		ACAGTTGGCT	CAGAAATGAG	AACTGGTCAG	GCCAGGCATG	GTGGCTTACA
		CAGCACTTTG	GGAGGCCGAA	GTGGGAGGGT	CACTTGAGGC	CAGGAGTTCA
45		AGGCAACAAA	GTGAGATACC	CCCTGACCCC	TTCTCTACAA	АААТАААТТТ
		CCAAATGTGG	TGGTGTATAC	TTACAGTCCC	AGCTACTCAG	GAGGCTGAGG
		GCTTGAGCCC	AGGAATTCAA	GGCTGCAGTG	AGCTATGATT	TCACCACTGC
50		GGGCAACAGA	GCGAGACCCT	GTCTCAAAGC	AAAAAGAAAA	AGAAACTAGA
		GTTTGTGGGA	GGAGGTCATC	ATCGTCTTTA	GCCGTGAATG	GTTATTATAG
	10876					

	AGGACAGAAA TTGACATTAG CCCAAAAAGC TTGTGGTCTT TGCTGGAACT CTACTTAATC
	10936 TTGAGCAAAT GTGGACACCA CTCAATGGGA GAGGAGAGAA GTAAGCTGTT TGATGTATAG
5	10996
-	GGGAAAACTA GAGGCCTGGA ACTGAATATG CATCCCATGA CAGGGAGAAT AGGAGATTCG
	GAGTTAAGAA GGAGAGGAGG TCAGTACTGC TGTTCAGAGA TTTTTTTTAT GTAACTCTTG
	AGAAGCAAAA CTACTTTTGT TCTGTTTGGT AATATACTTC AAAACAAACT TCATATATTC
10	11176 AAATTGTTCA TGTCCTGAAA TAATTAGGTA ATGTTTTTTT CTCTATAG GAA ATG AAT
	11233
	Glu Met Asn 85
15	CCT CCT GAT AAC ATC AAG GAT ACA AAA AGT GAC ATC ATA TTC TTT CAG
13	Pro Pro Asp Asn Ile Lys Asp Thr Lys Ser Asp Ile Ile Phe Phe Glu
	90 95 100 AGA AGT GTC CCA GGA CAT GAT AAT AAG ATG CAA TTT GAA TCT TCA TCA
	11329
20	Arg Ser Val Pro Gly His Asp Asn Lys Met Gln Phe Glu Ser Ser 105 110 115
	TAC GAA GGA TAC TTT CTA GCT TGT GAA AAA GAG AGA GAC CTT TTT AAA 11377
	Tyr Glu Gly Tyr Phe Leu Ala Cys Glu Lys Glu Arg Asp Leu Phe Lys
25	120 125 130 135 CTC ATT TTG AAA AAA GAG GAT GAA TTG GGG GAT AGA TCT ATA ATG TTC
	11425
	Leu Ile Leu Lys Lys Glu Asp Glu Leu Gly Asp Arg Ser Ile Met Phe 140 145 150
	ACT GTT CAA AAC GAA GAC TAGCTATTAA AATTTCATGC C 11464
30	Thr Val Gln Asn Glu Asp
	155
	(15) THEODMARTON FOR ORD TRUE
35	(15) INFORMATION FOR SEQ ID NO:14: (i)SEQUENCE CHARACTERISTICS:
	(A) LENGTH: 28994 base pairs
	(B)TYPE:nucleic acid (C)STRANDEDNESS:double
	(D)TOPOLOGY:linear (ii)MOLECULE TYPE:Genomic DNA
40	(vi)ORIGINAL SOURCE:
	(A)ORGANISM:human (F)TISSUE TYPE:placenta
	(iX) FEATURE:
	(A) NAME/KEY:5'UTR (B) LOCATION:115606
45	(C)IDENTIFICATION METHODS:E
	(A) NAME/KEY:leader peptide
	(B)LOCATION:1560715685 (C)IDENTIFICATION METHODS:S
	(A) NAME/KEY: intron
50	(B)LOCATION:1568617056 (C)IDENTIFICATION METHODS:E
	(A) NAME/KEY:leader peptide
	(B) LOCATION: 1705717068
	(C)IDENTIFICATION METHODS:S (A)NAME/KEY:intron
55	(B) LOCATION: 1706920451

```
(C) IDENTIFICATION METHODS: E
                  (A) NAME/KEY:leader peptide
                  (B) LOCATION: 20452..20468
                  (C) IDENTIFICATION METHODS:S
5
                  (A) NAME/KEY: mat peptide
                  (B) LOCATION: 20469...20586
                  (C) IDENTIFICATION METHODS:S
                  (A) NAME/KEY: intron
                  (B) LOCATION: 20587..21920
                  (C) IDENTIFICATION METHODS: E
10
                  (A) NAME/KEY: mat peptide
                  (B) LOCATION: 21921..22054
                  (C) IDENTIFICATION METHODS:S
                  (A) NAME/KEY: intron
                  (B) LCCATION: 22055..26827
15
                  (C) IDENTIFICATION METHODS: E
                  (A) NAME/KEY: mat peptide
                  (B) LOCATION: 26828..27046
                  (C) IDENTIFICATION METHODS:S
                  (A) NAME/KEY:3'UTR
                  (B) LOCATION: 27047..28994
20
                  (C) IDENTIFICATION METHODS: E
             (xi) SEQUENCE DESCRIPTION: SEO ID NO:14:
       ACTTGCCTTA AAAGCTTTGC ATAGGTAGAC AACATTAGAT TAATTTCCTT GCTCACATCT
       60
25
       GTTCAAGAAA AATCATTTAA GTTATAAAAT ATAACAAACC TTCTGCATTA TAAGACTGAT
       120
       GTTTAGAAAT ATAAACATTT TATACATCAC CATTTAAATC TTTCTCCAAG GCTTCATCTT
       180
       TATAAAATAG TCCGGAAATT TCAGAGAAAG ATGAATCTGA TTTTCCAAGA GAGGACAGCT
       240
30
       GTGGACTATC TGGCACTGGA GACTAAATAA AGAAAGCAGG TACAGTCAAT AAGATCTTCA
       300
       GGACATATAC ATTTTGTTTA TTAAGAAAAA GCAAATAAAA CATTTTTCAG AAAAAGGCAA
       360
       ACATGCTAGA AAGCATATGA CTTAGTCATT TGAGTTTTTA TTATTAAGGA AATTTACAGG
       420
35
       CCCAAGAAAC ACCTTGCTCA ATATATTAAA TTTTATTTTG GTTTTCAACT AGACTTTGCT
       TTTCATTTGT TTGTTTTTGT GACAAGTTCT CGCTCTGTCA CCTAGGCCAA AGTGTAGTGA
       540
       CACAATCTTA GCTCACTGTA GCCTCCTAGA TTCAAGTGAT CCTCCTGTCT CAGACTCCTG
40
       600
       660
       GTTTTCAGAG ACAATGTATT GCAGCGTTGC CCAGGCTGAT CTGAAACTCT TAGCCTCAAA
       720
       CGATACTCCT GCCTCAGCCT CCCAAAGCAC TAGGATTACA GACATGAGCC AATGCGCCCA
45
       780
       GCCTTAAATT AGACTTTAAA TGTGGTTTTA AACTCCTGTT GAAAAAGCGT CTGGTATCTT
       GAACCAGTAG ATGTTTTCAT AGCAATGAAG CTAAACTGTA ATTTAGACAG TAGCCAAATG
       900
50
       CTTGTGAAAT TTTGCTAAAT AATATAATCT TCAAGGGAGC AAATCATGTC CCAAATGCAA
       960
       AAGATCAACT GGTGGGGGCA GTAGTAAAAG ACAGGATACT GTGCTCTTTA AAAGGTCAGT
       1020
       AACTATAGTA CCTAGTTATC TTACTTATCA CAGCAAAATA ATTACATAAA ATCCTATGGA
55
       TCATAAAGGC ACAGACTCAC TTCTGTCTCT AGATCTCAAG CTACCAAAAA GAAATCTCCC
```

						•
	1140 AATAGTTTCT	TGGAGGCCTA	TACTTAGTGA	AAAAGCAGCT	GGAATCAACA	TAGTTCCTCC
	1200					
5	1260					CTTATCTATA
	GAGCTTTCTT 1320	AAAGGGAAGA	AATTTGAGTA	GTATGTAAAA	CAGAATAAAA	GATTAAGGCT
	CCATAGGCAT	ACAGCTTACC	TCCAATTCTC	TTGGCCTCTT	GCAATTTCTA	TTATCAGGCT
10		ATTTGCCATC	ATATTCCGAA	GGCACCAGCT	ACAAAGCTTA	GAACAATGCC
		ACAAACTCCA	TGCTACAAGC	TCTCTGGAAT	CCTTCCCTGT	TTCCCACTCC
		GTTAATTTAG	ACTGTCATTA	TCTGTCACTT	TCCTAAACTC	AATTTCTCCC
15		CATTCTATCA	ACTGCTATTT	GGGTAATCTT	TCAAAACTTT	GATTACTGCA
	TTCCTTTAAC	TCAAAAACTT	TCATTGTTCC	AGAATAAGTT	GAAATTCCAT	GATATGGCCT
	TCAAGGTCCT	GTATTATCTG	GTGCAAGCCT	ACTAGTCCCA	TCATTTTCAA	CTACTCCTCT
20		AGCCAAATGA	GTCTCTCTGG	CAATTCTGCC	TTGTTTCAGG	ACTGGCTCAG
	1800 TTAAGATTCT 1860	TTTATCTTCG	GCCGGGCGCG	CTGGCTCACG	GCTGTAATCC	CAGCACTTTG
25		GCAGGAAGAT	CACCTGAGGT	CGGGAGTTCG	AGACCAGCCT	GGCCAGCATG
25		GTGTCTACTA	AAAATCCAAA	CATTAGCCAG	GCGTGGTGGC	AGGCGCCTGT
		ACTTGGGAAG	CTGAGGTGAG	AGAATCGCTT	GAACCCAGGA	GAGGGAGGTT
30		GAGATTGTGC	CATTGCACTC	CAGCCTGGGC	AACAGAGCGA	GACTCCACCT
	CAAAAAAAA	AAGGATTCTT	CTATCTTCAC	AAAATCTTAA	TGTTTAAACA	GGTCTTACAG
	2160 TTCATCTAAT 2220	TCAATCTCAT	TTTTTACAAG	TGAGAAAACA	GGGACAGTGA	CGGTGGATCA
35		GTAAGACTGA	GCTAAATTAG	AACCGAGATC	TCACTCGAGT	CTGAGGTTAT
	TCCCACTGTC	CAACCTTACT	TTAAAGTAGC	TTCAAATTTT	ACTTTTACTT	TTCCATAAAT
		TTTTCCCTAG	GAGTCCAAAT	GTTGAAACCT	GGAAGGGTAT	AGTCTCTGTG
40		GAGGGGAGCC	CTGTCCATAT	TCAAGTTATC	AATTGACTTT	GTTGTTTTTG
		CTGATTTGGG	TAACTTTAAC	ACATCTGTTT	GATTAGTCCT	ATAAAATATG
		AGACAGAAAG	AGCAACAACA	AATTTGAAAG	ATGCTTGTTA	AGTAAATTCT
45	2580 GTATCGTACG 2640	TGTCCATTCC	TGCCAGTACC	TTTATAGTAT	GTAAGTTTAC	GTGCTGTAAT
		GTATCTAGAA	AATACTACAC	ATGCACAGCA	GTGCTAACTT	TGCCTTGGGA
		ACTTCAGAGA	AGCCAACAGG	CAGATTTTTC	TCTCTTCCCT	TCCCCTTCTA
50	ATTTTCCCTT	TCCCCTTCAC	ССССТТСТСТ	TCTCTCCCCA	AGTAACACTG	TGCACCTATG
	Z820 TCAAACGAAA 2880	ACTTATAATC	AAGTAACTGT	TTCTGCAAAA	ATAAGTTCGT	TTTCCTGTCA
55		CCTCAGCAGA	TCCAGGCCTG	GTGGACGGGC	TGGTCTTCGT	CGTGTGCCAA

	ACACTGACCA	CTGCCCTGGC	TCTGCCATCT	TAGGCTTAGT	GACCTGGCTG	TTACTAAGCA
	CTGTCCCCTC	TGCCCCATGC	AGCTGTCTCC	TTCTAGTCTT	CTCCCTCTTC	TCAACGCGAT
5	CCTAGCCCCT 3120	CAGGCCATTT	CACCTCCATT	TTCCCTCACT	TCCCGCCGCC	CCTCCGCACT
	3180				•	CATCCTCGCA
10	CCCTTCCTTG 3240	TGTCACAGCC	CGTCACATTC	TCACAGGCGC	CCATCCCTCC	AGCCCCACCC
70	CAAGGCCAAT 3300	GTACTTCGCG	GTATGGGGAC	CTTCCTCGTC	AGCGAACGCG	AGGGAGTGAA
	GACCCTGGGC 3360	GCGGGGTGCT	CGGACTTCGG	GGGTGGAGGT	GGGAAGCGCG	CCGCACTCCC
15	AGCAGCCCCT 3420	GCACGAGTCA	CGTGACAGCT	CTCCCACCAC	CACCCCCCC	AACTTCCCCA
	CCGTAGCCTC 3480	CCAGAGCCAG	GCCCCACGGA	AAGGCAGCTT	TTTCCCGGTT	TTCTCCCGCT
	CTTTCCCCTC 3540	CACTTGGAAT	ACTCGTGAAA	CAAAAATCTC	TCCCTGCCAC	CCTGTGTGTG
20	TTTGAACCAG 3600	GAAAAAATCT	GAAACTGGTC	AAGAAAGAAC	AAGGAAGACT	TGCCAAAGCA
	AGGCCGGTGT 3660	GTGTCCCAGC	AGCTTAGAAT	CTCAGCAAAG	GAACACAAAA	TAGCACATCC
	ACGGCCTCTT 3720	TTCGAGTAAA	ATTTACTTGG	TTTGTTTGCA	GGAAGGGTTT	AAAACTGCGT
25	TTGCAGATGC 3780	TCTGTTTGCA	GGAAGGCTTT	AATCACGTGT	TCCCCTGGCC	CACAAGCAAG
	GCTTTTAGAT 3840	CCAGAGCCTC	AGTTACTGCC	CCCTCTTCCT	CTTTGGTGCA	ACCAAACGTT
	CAGAATCACG 3900	CCTTCTTAGA	AAATTCTTAC	CCCGGGTGTG	TCAATAAGTT	AAGTCTAATT
30		ATCAAAAAGT	GTTGCATAAC	ACACATGGCT	CACATAATTG	TAGCTTTGCC
	TCATCGGGTG 4020	TTTTAATGCG	GAGGCTTTGA	CCTGCAATTT	CAAAGATATA	CATTCCAAGC
35		TTAGTGGATG	TGGAAGAAAA	AAAAAAGCAA	ATTACCTCAT	AACACAAAGG
33	TCAATAACAC 4140	ACATCCATAA	GCTCCAGGTA	CAAAATCTTA	CATCTTAGAG	AACTATATTT
	•	TACATTACTA	AGGTTTTTTT	TTTCCTTTTG	CTTGATTAAA	TGTTAGTTAT
40		TGGAATTATT	CTGTGTGTGT	ATATTTATTT	GCTGTTTGTG	AAGAAGCCGG
		TAAGTTCCTA	GAAAATAAGC	GCTCAATGTG	TTTAATCTGA	GTTGCTAATA
		TAGGCCACAT	AATACTAGCC	TAGATAACTA	TGGCGAAGTA	AGGAGTCTCA
45		CAGAACAATA	GCAATCTGTG	TTGAATTTTT	ACCCTCTGTG	GTAAAATGAA
		ATGAAGTTTT	AGTTTGCCTT	AATTTTTATC	TTTATTGTTT	CAGACTCTTC
		AGTTTTCATC	AAGTCAAATA	TATTCACTTT	AAAGTGACTG	TGCTTTATTC
50		TCCTTCCTAA	TTTGGGGGGC	CAGGTGAGAT	AAGTTTTATG	AAATAAAAAG
		CTTACATTTT	TAGTGTCCTT	CCTTGGTAAA	ATGTAGAGTT	GTCCACTGTG
		CCTCCTTATT	ATCATGGTTG	CTGTTATTAT	TTTTAATGGT	TCATTAAACC
55		GGAAATACTC	ATGGAATTCA	TCTCACAGCC	TTCACACTGT	ATGATATTTA

	4800					
		TGTCCATCTG	ATTCTTAAAA	TATTTCCAAG	AAAAATGATT	CCACCTAAT
	4860 CATAAATGC1	TTCATCAGAT	TAAGAGAACA	CCATGGACAT	* TTTATTTTAT	' ጥጥጥ <b>አ</b> ጥጥጥጥጥ'
5	4920					
	AAATATTAAC 4980	TTCCATTGCA	TAAGCTAAAT	GGGTAGGAAT	AAGTGAGATG	ATATTGTTA'
	CTAGAGCTTT	AAAATATTCA	AAGGGCTGTC	ATCATTATCT	CATTTAATCT	TTGAAAACA
	5040	TACAAAGGAC	· ACTCACACAT	·	ስጥስጥ <u>ሮ</u> ስ ስ ስ <mark>ሮ</mark> ለ	3 3 3 3 CTCT'
10	5100					
	TGTCCCAAAA 5160	CTTCAAAATG	TGTAAATTAC	ACATTCTGCA	TCTTTACAGC	TGGAGAAAA
		TGGAATATTT	AAAATTAGAG	CTTGCTTAGT	GTGCTGCTTC	TGATCACTA
15	TTGATCCCAC	TTCGTGCTTT	CATGTTAATT	GGCCCAATTG	GACTCTACAG	TTGGAAGGT
	5280 AAAACTTACT	ATTTCAACTT	GAGTCACGTA	<b>ፐርፐል</b> ፐፐርፒፑል	TCATATACTT	CTT A A A CCTT:
	5340					
	CTATTTTTT 5400	TCTTCTGATA	GTCACCACAC	CAAGCACTTC	CAGCCACCCT	GCCACAGACT
20	TCCTTTGTAA	TCACTGTTGA	AGGACATGAT	GTTTTTATGA	CTTCCCGAAA	TGAAAACCCT
		TAAAACAAAC	AAACCAACAA	AAAGTAGTGT	TTATGTAAGC	ATTTTGTTC
	CTGACTCTAG	GAACCCCTCT	GTTTTTATAT	CAACTCTGTA	CTGGCAAAAC	ACAAAAACA
25	5580 AATGCCACCT	TGCTAATTCC	CTTCCTAGCA	ΑΔGΤΑΔΤΔCΔ	GTTTAGCACA	TGTTCN AGAI
	5640					
	5700	AAGAAATTTT	GTTTCCACTA	ATTATTTTCA	AGACTGTGAT	ATTTACACTO
30	5760	ACGTTACATT				
	ATATATGTAT 5820	CCTTAAATTG	TATTTCAAAT	ATTTTAGGTC	AGTCTTTGCT	ATCATTCCAG
		TTTTAACACT	GGAAACTGCA	AGTAAATATT	TGCCCTCTTA	CCTGAATTTI
35	GGTAGCCCTC	TCCCCAAGCT	TACTTTCTGT	TGCAGAAAGT	GTAAAAATTA	TTACATAAA
		GGTATCCGTG	TGGCTTGCAT	CTGATACAGC	AGATAAAGAA	GTTTTATGA#
	6000	TGTTCCACTG	~~~~~~~~~~	CEETA A MCCCCC	MCM > MC > > CM	> m.c.comma.>.c
	6060					
40	ACCATATTGA 6120	GCTTGGGAGG	AAGGGGAAGT	CCTGAATGAG	GTTATAAAGT	AAAAGAAAAT
	ATTTGCAAAA	TGTTCCTTTT	TTTAAAATGT	TACATTTTAG	AAATATTTTA	AGTGTTGTA#
	6180 CATTGTAGGA	ATTACCCCAA	ТАССАСТСАТ	таттссссат	тсталллтлл	C
45	6240					
	6300	TGTGACCAGG	AAGTCTGAAA	ATGAAGAGAG	ACAGATGACA	AAAGAAGATG
	CTTCTAATGG 6360	ACTAAGGAGG	TGCTTTCTTA	AAGTCAGAAA	GAGATACTCA	GAAAGAGGTA
	CAGGTTTTGG	AAGGCACAGA	GCCCCAACTT	TTACGGAAGA	AAAGATTTCA	TGAAAATAGI
50	6420 GATATTACAT	TAAAAGAAGT	ACTCGTATCC	<b>ፐርፕር</b> ርርልር <b>ጥ</b> ጥ	<b>ቸል</b> ጥጥጥር ልርጥ	тесаттесе
	6480					
	TAGGAAAGAG 6540	CCTGTTTGAA	GGCGGGCCCA	AGGAGTGCCG	ACAGCAGTCT	CCTCCCTCCA
55		CATTCTCTCC	CCAGCTTGCT	GAGCCCTTTG	CTCCCCTGGC	GACTGCCTGG
33	6600					

	6660					ACTCTGGCTG
	CTAAAGCGGC 6720	TGCCACCTGC	TGCAGTCTAC	ACAGCTTCGG	GAAGAGGAAA	GGAACCTCAG
5		TCGCTTCCTC	TCGCAACAAA	CTATTTGTCG	CAGGTAAGAA	ATATCATTCC
	TCTTTATTTG 6840	GAAAGTCAGC	CATGGCAATT	AGAGGTAAAT	AAGCTAGAAA	GCAATTGAGA
10		CCATCTAGCA	TCACTACGAT	GAGCAGTCAG	TATCAACATA	AGAAATATAA
		AGTAGAATTT	TTTTCTTTTA	TCAGATATGG	GAGAGTATCA	CTTTAGAGGA
		AACTTTTTGC	TCTCATGTTC	CCTTTACACT	AAGCACATCA	CATGTTAGCA
15		TTTTAATTAA	AAATAACTAT	GTACTTTTTT	AACAACAAAA	AAAAGCATAA
		TTTTTTTTTT	TTACAAGTGT	TTTAACTGGT	TTAATAGAAG	CCATATAGAT
		CTCATCTGCT	TTGCATTCAG	ACTACTGCAA	TATTGCACAG	AATGCAGCCT
20		CTGTTGTACA	CTCATGAGAG	AATGGGTGAA	AAAGACAAAT	TACGTCTTAG
		AATAGCTTTC	ACTTTAGGAA	CTCCCTGAGA	ATTGCTGCTT	TAGAGTGGTA
		GCTTCTCTTT	AAACGGAATC	TCAAGACAGA	ATCAGTTACA	TTAAAAGCAA
25		TGCCCATGGT	TAGTCATCTT	GTGAAATCTG	CCACACCTTT	GGACTGGGCT
		AATATAGCAT	TCCCCGAGAT	AATTTTCTCT	CACAATTAAG	GAAAGGGCTG
		TCTGTTTGAA	GTTGAATAAC	AAAAATTAGG	ACCCCCTAAA	TTTTAGGGCT
30		GTCTTTTTGC	CTATATTCAG	CTACTTTACG	TTCTATTAAA	TCTTCTTTCA
		ACTAGCTCAT	GCCTAGAATC	TCAGGCAGGC	CTGAGCCCAG	GAATTTGAGA
35		CAACACAGTC	TCTACAAAAA	AATAAAAAAT	TACCTGGGTG	TGTTGGTGCA
		ACTACTCAGG	ATGCTGAGGA	CTGCTTGAGC	CCAGGATAGC	CAAATCTGTG
		CCACTAAACA	GAGCGAGACT	TTCTCAAAAA	AACAAACAAA	AAAACAAACA
40	. •	AAAATAACTT	TTTATCTGCA	ATGTTTTCCT	ATTGCCTGTG	AGATTAAATT
	-	CCTGATTTCC	AAAGCCCTCC	ATAATCTAAT	CCGACTTTAC	CTTGTGTTCA
		GCAGGACTGT	TCCACTACAA	TCCAAAAATC	ACAGGTTGGG	TGCAGTGGCT
45		ATCCCAACAC	TTTGGAAGGC	CAAGGCAGGT	GGATTGCTTC	AGCTCAGGAG
		GCCTGGGCAA	CATGGCAAAA	ACCCTGTCTC	TCCAAAACAT	ACAAAAATTA
		GTAGTATGTG	CCTGTAGTCC	CAACTACTCA	AAAGGCTAAG	GCAAGAGGAT
50		CAGGAGGTCA	AGGCTACAGT	GAGCCATGTT	TACTGTGTCA	CTGCACTCCA
		TAGAGCAAGA	CCATGTCTCA	ааааааааа	AAAGAAAAGA	AAAGAAAAA
		ATTCAGTTCA	CCCCCACCAC	AACATTGTTT	TGATTATCAC	ATAAATGCTG
<b>55</b>	GTCCATTGCC	TTCTCTATCT	ATTCAAATCT	TTAAGCATTC	TTTGAGATTC	AACTCAATTC

	8460					
	TCCTTTTCAA 8520	ACTAGGCCAT	TTAAACTACA	TCAGTTCCAT	TTTGATTTTC	TTGCTTTGA
5	TCTACAGACT	САААААСААА	AACTTAAAA	CTTATTTTT	AAGTTTTCTG	CTACTCTCA
	TTCTTCAACA	CTCACATACA	CGCATTCATA	ATAAGATGGC	AGAATGTTCA	AGGATAAAA
	GATTTATAGA 8700	ACTGAAAAGT	TAGGTTTTGA	TCTTGTTGCT	GTCAAGATGA	CTACCTACC
10	GATCTCAGGT 8760	AATTAATTAT	GTAGCATGCT	CCCTCATTTC	ATCCCATACC	TATTCAACA
	GATTGGAATT 8820	CCACAGCAAG	GATAAACATA	ATCATAGTTG	CTTTTCAAGT	TCAAGGCAT'
	8880	ATCTAGTAGT				
15	8940	AGGCTGGAGT				
	9000	AGTTATTCTG				
20	9060	GCTAATTTTT				
20	9120	ACTCCTGACC				
	9180	AAGCCACCGT				
25	9240	TTGTATAATA				
	9300	GGCCGTTTGC AGGTTGGTAG				
	9360	GGGAACTGAG				
30	9420	CTAATTTCCA				
	9480	ATAACTAAAT				
05	9540	TCTTTCTGAG				
35	9600	ACGAAATAGT				
	9660	TCAAAGAAAC				
40	9720 TGGGGATGAG	AACACACTAC	TTGTAATCAG	TCATAGATGT	ACTGAGAACT	AACAAGAAG <i>i</i>
		AATAAGAATG	AAGAATTCAA	AATCAACACA	TGAAATAAAA	AGAAACTAC1
		GGAGAAGACA	TTAGAAAAAT	TATTCTATTT	ТТААААТТСТ	GTTTTCAGGC
45		ттсттсстсс	TTCTCATTGG	TTTTCAGGTG	GAGGGAAAGT	TTAAGATGG#
	9960 AAAAATATAT 10020	ATATTCTACA	CATCCCTTTC	TACGCTGTTG	TCATGGCAAC	AAGGTTTATC
		TTTATTCATA	CAACATTTAT	TGAGTTCTTA	CTGTGTGGTA	AGCTCTTTCC
50		AATTCAGGGG	AAAAAAGACA	ACTCATTGTC	TTAAAACTCA	GATGAAAGCI
		ATTTTTAATC	AAAGTAATCT	CAATTTAGGG	TAGTAAGAGC	TATTTAAGAA
55		GTGTGAAGGA	GGTAGGACTC	TGAGGAGAGA	ATAGTTAGCT	AGGAATGAAA

	GAGCAGAGAA	GTTTTCCTAG	AGGAACTATT	AAAGCTGGGA	GTTACGGGAT	GAAAGATGAG
	10320 GCAGGGTTTG	CAGGCAAAAA	ממממממממ	GCAGGGGAAG	GGGAAGTTCT	GCCCTCCCAG
5	10380	CCOCIBBBB.		GCAGGGGAAG	GGGAAGIICI	GGCCIGGCAG
3	10440				AAGAAAACCA	
	TATTAAAATA 10500	GAAGATGCCA	GGGGTAATGA	GGGCTTGATT	TAAAACAGTG	CTGTTGGAGA
10	10560				GAACCTACAG	
	ACAAGGGAAA 10620	GATTAGACAA	AGGAGTTAAG	AATGACTCCC	AGGTTTCAGT	TTGGGGCAGG
	TAACTAGGAC 10680	ATGTTTTGAA	AAGTAATGTA	TTGGATCTCT	TACCATTGGA	ACTATGTATG
15	10740				CCCACCACCA	
	10800				AGTTTGTAGT	
	10860				ACCATCTGCT	
20	10920				CCAAATTCAA	
	TAAGCTTATA 10980	TGTAAATTAG	GAGCTCTACA	GTTTGATTTC	GAGCAGCCCC	TCCTGAAACC
<u></u>	CTTTCTCTTT 11040	CGACTTCTGT	GACACATCTC	AGATTTACAA	AACTGAACTA	ATTATTTTAC
25	ACTTGAGCTG 11100	TATTTTCGTT	CTTCTTTCTT	GATGAATGAG	GTAACCACTC	AACAAATTGC
	11160				TTCTGTTTGA	
30	ATCAGCTGAG 11220	AAATCCCGCT	GTTTAGTATC	TCTTGAATTC	ATTACCTTAA	TTTATAGCCT
	11280				TCTGACCTCT	
	GTTCAATCAG 11340	GTCCATTCTT	TTGTTCTTGG	TGGTGGTGGT	GGTGTTGACA	GAGTTTCGCT
35	TTTGCTGCCC	AGGCTGAAGT	GCAGTGGAGC	ACTTCACTGC	AACCACAGCC	TCCTGGGTTT
	11460				TGCCACCACA	
	TTTGTGTTTT 11520	CAGTAGAGAC	AGGGTTTCAC	CATGTTGGTC	AGGCTGGTCT	CAAACTCCTG
40	ACCTCAAGCA 11580	ATCCACCCAC	CTCAGCCTCC	CAAAGTGCTG	GGATTACAGG	CATGAGCCAC
	11640				TGAGTTTTTA	
	11700				TTCAGGAAAA	
45	11760					AGCTTTCATC
	11820				GCAACGGTAT	
50	11880				ATAGTTTATC	
50	TAATTTCAAA 11940	GCTGTCAGGT	CAAATGAGTT	ATAAACTGTT	AACACTATTG	CCACATGCAA
	12000				TTATTATGTA	
<i>EE</i>	TAAATAACTC 12060	AGAAGTTCTT	CAGACATACA	GGTTATTATT	GTGCTTTTTA	AACATAATTT
55		TATATATGAT	AATGTTATCC	AAGTGCTAAG	GGATGTATTG	TTACTGCTGT

	12120					
	GCAAAAAAAA 12180	AAAAAAAAA	AACTCCAAAT	AAATATGTTG	AAACCAAGTT	TATATGCAA
5		TAAAAAGGCC	AAAGTACCAC	CATAATAGGC	TGTGTGGAGA	CGGCAGGCT
		GTAATAATGC	: TGAGAAAGTT	GAAAAAAGAA	AGAAAGCAAC	AATATGCTT
	GGTTGTTGTA	GGTTTATGTA	CTCCAAGAAT	ATCTCCTCTC	AAACTTTTAC	GTTTTTTCC
10		' AACTTTGGCT	GGGCGCAGTG	GCTCTTGCCT	GTAGTCCCAG	CCTTTGGGA
	GCCAAGGCGG 12480	GCAGATCACC	TGAGGTCAGG	AGTTTGAGAC	CAGCCTGACC	AAAAATGGA
	AAACCCGCCC 12540	CCCTCACTAC	TAAAAGAATA	CAAAATTAGG	CCGGGCACAG	TGGCTTACC
15	12600				ACCTGAGGTC	
	12660		•		ATTAGCCGGG	
20	12720				AATCACTTGA	
20	12780				GCCTGGGCAA	
	12840				TGAACTATGT	
25	12900				ATTTATGGTA	
	12960				TAGCCTTTGT	
	13020				TGCTTCTCTT	
30	13080				AAACAGAAAG	
	13140				TAACTCTTTA	
	13200				CTCTCATTCA	
35	13260				CAAACCTTAA	
	13320				ACTAACTATG	
	13380				CTTGGTACAT	
40	13440				ATGTACAGAA CTCAGGTTAA	
	13500				GCTGATAATA	
45	13560				GTGGAGCTCA	
	13620				TTGGTGAAAA	
	13680				CTCTAAGAGT	
50	13740				TTATAAATTT	
	13800				TATTTAAATT	
	13860				TTATTTAAAT	
55	13920				**************************************	0.0110000

		CTGACACCTA	TAATCCCAGA	ACTTTGAGAG	GCCAAGTCAG	GCAAACCATT
	13980	) COOTTO ) C ) C	61.666ma66a	) ) COMO COMO )		4
	14040	AGITIGAGAC	CACCCTGGGC	AACGTGGTGA	AACCCTGTCT	CTACCAAACA
5		TATCTGGGTG	TGGTGGCACG	CATCTGTGGT	CCCAGATGGG	AGTCCCAGGC
	TAAGATGGGA 14160	GAATCGCTTG	AACCCAGGTG	AGAGGGGTGG	GGTGGATGTT	GCAGTGAGCT
10	14220	CACTGCACTC				
	14280	TAAATAAGAT				
	14340	TAAAACTATC				
15	14400	GAGGCCAAAA				
	14460	CTTAATTCAG				
	14520	TGCTAGGCCT				
20	14580	ATACCATTTA				
	14640	TTAAATAATA TAAACCATGT				
	14700					
25	14760	TAATGACAGA				
	14820	GGGAGGAATG				
30	14880	AGTTCACAGA				
	14940	CACCGTATTG				
	15000	ATCCTGTATC				
35	15060	TAGTATAATC				
	15120	AATGCCAGAG				
	15180	TTATCATTGT				
40	15240	CATAGGAACC				
	15300	AGATGTGTTT				
	15360	CTCAAGGAAT				
45	15420	ATATTGTTTA				
	15480	ACTTAAACAT				
	CAAGATTGTT 15540	CATCAGTATA	CTAGGTGATT	AACTGACCAC	TGAACTTGAA	GGTAGTATAA
50	AGTAGTAGTA 15600	AAAGGTACAA	TCATTGTCTC	TTAACAGATG	GCTCTTTGCT	TTCATTAGGA
	ATAAAG ATG 15651	GCT GCT GAA	CCA GTA GA	A GAC AAT 1	GC ATC AAC	TTT GTG GCA
	Met	Ala Ala Glu	Pro Val Gl 3-	-	•	Phe Val Ala
55	ATG AAA TTI	ATT GAC A	-	_	-25 G GTAAGGC	TAATGCCATA

	15702					
	Met Lys P -20	he Ile Asp	Asn Thr Leu	Tyr Phe Ile	Ala . -10	
5	GAACAAATA 15762	C CAGGTTCA	GA TAAATCTAT	T CAATTAGAAA		AGGTGAACTA
		T CTTTGTGTG	CA CCAAATTTC	A CTGTAATATT	AATGGCTCTT	AAAAAAATAG
		A GAAATTAA	CC ACAACATGT	C CAAGGTCTCA	GCACCTTGTC	ACACCACGTO
10		T TTAATCAGO	CA GTAGCTCAC	r ctccagttgg	CAGTAAGTGC	ACATCATGAA
		T TTCATGGG	A AATCCCAGT	T TTCATTGGAT	TTCCATGGGA	AAAATCCCAG
		G GGTGCATTO	A GGAAATACA	A TTTCCCAAAG	CAAATTGGCA	AATTATGTAA
15		T AAATTTAGA	G TTCCGTGAA	TACACCATTT	TATGTAAATA	TGTTTGACAA
		G ATTCTTTT	T TTTTTTTCT	G TTGCCCAGGC	TGGAGTGCAG	TGGCACAATO
		T GCAACCTCC	A CCTCCTGGG	TCAAGCAATT	CTCCTGCCTC	AGCCTTCTGA
20		A CTACAGGTG	C ATCCCGCCA1	GCCTGGCTAA	TTTTTGGGTA	TTTTTACTAG
		T TTGGCATGT	T GTCCAGGCTC	GTCTTGGACT	CCTGATCTCA	GATGATCCTC
25		G CTCCCAAAG	T GCTGGGATTA	CAGGCATGAA	CCACCACACA	TGGCCTAAAA
		T ATGATTAAT	C TCCTGTGAAC	AATTTGGCTT	CATTTGAAAG	TTTGCCTTCA
		r tcatttaaa	A GCCTGAGCA	CAAAGTGAGA	CCCCATCTCT	ACAAAAAACT
30		C CTGTGGACA	C CTCCTACCTT	CTGTGGAGGC	TGAAGCAGGA	GGATCACTTG
		a TTTGÄGCCT	G CAGTGAGCTA	TGATCCCACC	CCTACACTCC	AGCCTGCATG
		C CTGACACAC	A CACACAAAAA	AAAACCTTCA	ТАААААТТА	TTAGTTGACT
35		r GACTTTCCG	T TTAAGCAATA	AATTTAAAAG	TAAAATCTCT	AATTTTAGAA
		TAGTTACAT	A TTGAAATTTT	TAAACCCTAG	GTTTAAGTTT	TATGTCTAAA
		A ACACACTAA	G TCTGATAAGC	TTCATTTTAT	GGGCCTTTTG	GATGATTATA
40	TAATATTCTC	G ATGAAAGCC	A AGACAGACCC	TTAAACCATA	AAAATAGGAG	TTCGAGAAAG
		A AAAGTAAAA	G CTAGAATGAG	ATTGAATTCT	GAGTCGAAAT	ACAAAATTTT
	ACATATTCTG	TTTCTCTCT	r tttccccctc	TTAG CT GA	A GAT GAT G	GTAAAGT
45				Ala Gl -10	u Asp Asp G	lu
	AGAAATGAAT	TTATTTTC	r ttgcaaacta	AGTATCTGCT	TGAGACACAT	CTATCTCACC
50	ATTGTCAGCT	GAGGAAAAA	A AAAAATGGTT	CTCATGCTAC	CAATCTGCCT	TCAAAGAAAT
		TAGCACAGC	r ttggaatgaa	GATGATCATA	AGAGATACAA	AGAAGAACCT
	CTAGCAAAAG 17315	ATGCTTCTC	r atgccttaaa	AAATTCTCCA	GCTCTTAGAA	TCTACAAAAT
55	AGACTTTGCC	TGTTTCATT	G GTCCTAAGAT	TAGCATGAAG	CCATGGATTC	TGTTGTAGGG

	GGAGCGTTGC 17435	ATAGGAAAAA	GGGATTGAAG	CATTAGAATT	GTCCAAAATC	AGTAACACCT
		ATGCTTTGGG	AAGAAGCCTG	GAAGGTTCCG	GGTTGGTGGT	GGGGTGGGC
5		GGAAGTAGAG	GAGATAGGAA	TGGGTGGGGC	AAGAAGACCA	CATTCAGAGG
		AAAGAAACCA	TGGCATTTAT	GATGAATTCA	GGGTAATTCA	GAATGGAAGT
10	AGAGTAGGAG 17675	TAGGAGACTG	GTGAGAGGAG	CTAGAGTGAT	AAACAGGGTG	TAGAGCAAGA
	CGTTCTCTCA 17735	CCCCAAGATG	TGAAATTTGG	ACTTTATCTT	GGAGATAATA	GGGTTAATTA
	AGCACAATAT 17795	GTATTAGCTA	GGGTAAAGAT	TAGTTTGTTG	TAACAAAGAC	ATCCAAAGAT
15	17855					GGCAGCTCAG
	17915	GCTGGAAGCA				
	17975					TACTTCCTGG
20	18035					TTTCAAAGAC
	18095	AGTTAAACAC				
	18155	CAAGTGCAAG				
25	18215	AATTCAGAAA				
	18275	GTGGGGTGAG				
30	18335	TGATGGTGTC				
	19395	TTTAGAAACA				
	19455	GAAAGTAGAA				
35	18515	TAGGGAACAC				
	18575	TTTCGTTTTG				
	18635	TACATGTGGT				
40	18695	ATACATAGAA				
	18755	AAGTGAGAGA				
	18815	GTAGAAAGAA				
45	18875	AGAGAATTCC				
	18935	AATTTTGCAG				
	CAACAAGGAG 18995	TTTGGTGATC	TCAGTGAAAG	CAGCTTGATG	GTGAAATGGA	GGCAGAGGCA
50	GATTGCAATG 19055	AGTGAAACAG	TGAATGGGAA	GTGAAGAAAT	GATACAGATA	ATTCTTGCTA
	19115	TGTTAAAAGG				
55	ATGGAGCAGT 19175	TTTAAATCTC	AAAATAAAGA	GCTTTGTGCT	TTTTTGATTA	TGAAAATAAT
55	GTGTTAATTG	TAACTAATTG	AGGCAATGAA	AAAAGATAAT	AATATGAAAG	TATAAAATAT

	19235
	AAAAACCACC CAGAAATAAT GATAGCTACC ATTTTGATAC AATATTTCTA CACTCCTTTC
	19295 TATGTATATA TACAGACACA GAAATGCTTA TATTTTTATT AAAAGGGATT GTACTATACC
5	19355
	TAAGCTGCTT TTTCTAGTTA GTGATATATA TGGACATCTC TCCATGGCAA CGAGTAATTG
	CASTTATATT AAGTTCATGA TATTTCACAA TAAGGGCATA TCTTTGCCCT TTTTATTTAA
10	19475 TCAATTCTTA ATTGGTGAAT GTTTGTTTCC AGTTTGTTGT TGTTATTAAC AATGTTCCCA
	19535 TAAGCATTCC TGTACACCAA TGTTCACACA TTTGTCTGAT TTTTTCTTCA GGATAAAACC
	19595
	CAGGAGGTAG AATTGCTGGG TTGATAGAAG AGAAAGGATG ATTGCCAAAT TAAAGCTTCA 19655
15	GTAGAGGGTA CATGCCGAGC ACAAATGGGA TCAGCCCTAG ATACCAGAAA TGGCACTTTC
	TCATTTCCCC TTGGGACAAA AGGGAGAGAG GCAATAACTG TGCTGCCAGA GTTAAATTTG
	19775 TACGTGGAGT AGCAGGAAAT CATTTGCTGA AAATGAAAAC AGAGATGATG TTGTAGAGGT
20	19835 CCTGAAGAGA GCAAAGAAAA TTTGAAATTG CGGCTATCAG CTATGGAAGA GAGTGCTGAA
	19895
	CTGGAAAACA AAAGAAGTAT TGACAATTGG TATGCTTGTA ATGGCACCGA TTTGAACGCT 19955
25	TGTGCCATTG TTCACCAGCA GCCACTCAGCA GCCAAGTTTG GAGTTTTGTA GCAGAAAGAC 20015
23	AAATAAGTTA GGGATTTAAT ATCCTGGCCA AATGGTAGAC AAAATGAACT CTGAGATCCA
	20075 GCTGCACAGG GAAGGAAGGG AAGACGGGAA GAGGTTAGAT AGGAAATACA AGAGTCAGGA
	20135 GACTGGAAGA TGTTGTGATA TTTAAGAACA CATAGAGTTG GAGTAAAAGT GTAAGAAAAC
30	20195
	TAGAAGGGTA AGAGACCGGT CAGAAAGTAG GCTATTTGAA GTTAACACTT CAGAGGCAGA 20255
	GTAGTTCTGA ATGGTAACAA GAAATTGAGT GTGCCTTTGA GAGTAGGTTA AAAAACAATA 20315
35	GGCAACTTTA TTGTAGCTAC TTCTGGAACA GAAGATTGTC ATTAATAGTT TTAGAAAACT
	20375 AAAATATATA GCATACTTAT TTGTCAATTA ACAAAGAAAC TATGTATTTT TAAATGAGAT
	20435 TTAATGTTTA TTGTAG AA AAC CTG GAA TCA GAT TAC TTT GGC AAG CTT GAA
	20486
40	Glu Asn Leu Glu Ser Asp Tyr Phe Gly Lys Leu Glu -5 1 5
	TCT AAA TTA TCA GTC ATA AGA AAT TTG AAT GAC CAA GTT CTC TTC ATT 20534
	Ser Lys Leu Ser Val Ile Arg Asn Leu Asn Asp Gln Val Leu Phe Ile
45	10 15 20 GAC CAA GGA AAT CGG CCT CTA TTT GAA GAT ATG ACT GAT TCT GAC TGT
	20582 Asp Gln Gly Asn Arg Pro Leu Phe Glu Asp Met Thr Asp Ser Asp Cys
	25 30 35
50	AGA G GT ATTTTTTTA ATTCGCAAAC ATAGAAATGA CTAGCTACTT CTTCCCATTC 20638
	Arg Asp
	TGTTTTACTG CTTACATTGT TCCGTGCTAG TCCCAATCCT CAGATGAAAA GTCACAGGAG
	20698 TGACAATAAT TTCACTTACA GGAAACTTTA TAAGGCATCC ACGTTTTTTA GTTGGGGTAA
55	20758

	AAAATTGGAT Z	ACAATAAGAC	ATTGCTAGGG	GTCATGCCTC	TCTGAGCCTG	CCTTTGAATC
	ACCAATCCCT 1	TTATTGTGAT	TGCATTAACT	GTTTAAAACC	TCTATAGTTG	GATGCTTAAT
5	CCCTGCTTGT T	TACAGCTGAA	AATGCTGATA	GTTTACCAGG	TGTGGTGGCA	TCTATCTGTA
	ATCCTAGCTA (	CTTGGGAGGC	TCAAGCAGGA	GGATTGCTTG	AGGCCAGGAC	TTTGAGGCTG
10	TAGTACACTG 3					
10	CCTTGTCTCT A					
	TCTACTGTGC (					
15	GAATGTGCAT 1 21238					
	TTAGCATTTA A					
	GTTGCAGCGT C					
20	CAGTTTCACT C			•		
	ACTACAACCT C					
0	GGATTACAGG C					
25	TTTCACCATG T 21598 GCCTCCCAAA C					
	21658 CTTTTATGAG T					
30	21718 GGATTACAGG C					
	21778 GCTGTACATA G					
	21838 CTATCCTTTG G					
35	21898 TCATTTATTA T				ACC ATA TT	
	21949			Ala Pro Arg	Thr Ile Ph	e Ile Ile
	AGT ATG TAT	AAA GAT AG	40 C CAG CCT A	GA GGT ATG	45 GCT GTA ACT	ATC TCT
40	21997 Ser Met Tyr		r Gln Pro A		Ala Val Thr	
	GTG AAG TGT 22045	55 GAG AAA AT	r TCA ACT C	TC TCC TGT	GAG AAC AAA	65 ATT ATT
45	Val Lys Cys	Glu Lys Ile	e Ser Thr L	eu Ser Cys	Glu Asn Lys	
	TCC TTT AAG 22103		AGCCTTACTT		CATGTTAATA	80 TAATCAATAT
	Ser Phe Lys AATTAGAAAT A	TAACATTAT '	ГТСТААТСТТ	מדמדממנדמ	ልፐርጥል ልጥጥልር	<u>አ</u> አአርጥሮአአአ
50	22163 TATCCTCAGA C					
	22223 TGAATACTTA C					
	22283 GCCTGTCACA G					
55	22323					· · · · · ·

	TCAGTCTTTA	TACAAATAAT	AATGTAGAAT	ACATATGTGA	GTTATACATT	TAAGAATAAC
		CCAGAATGAG	TTCTGCTATG	AAGAATGAAG	СТААТТАТСС	TTCTATATTT
5		GTAAATTATG	ATAATATTT	AATCCCTAGT	TGTTTTGTTG	CTGATCCTTA
		TAGACACAAG	CTTCAGCTTC	CAGTTGATGT	ATGTTATTTT	TAATGTTAAT
10		AAAAGTTATG	AGATCAGCTG	TAAAAGTAAT	GCTATAATTA	TCTTCAAGCC
10	AGGTATAAAG 22703	TATTTCTGGC	CTCTACTTTT	TCTCTATTAT	TCTCCATTAT	TATTCTCTAT
	22763			GATAAACCAC		
15	22823			CAAATTGGCA		
	22883			TTGCCAATAA		
	22943			GACCTTAGTG		
20	23003			AATCCAAGTA		
	23063			TCTCCCTCCA		
	23123			TGACAGGCAG		
2,5	23183			AAGGGTTAAG		
	23243			AAGCTATGTG . CCGAACCTAC		
30	23303			TAATGCTTGG		
	23363			TCAATTTTGA		
	23423			TAATCCCAGC		
35	23483			CAGCCTGACC		
	23543			GTGGCATATG		
	23603			CGGGAGGCAG		
40	23663			GAGCAAAACT		
	23723	•		таатааттта		
	23783			GAGGGAATCT		
45	23843			CAATAAATAT		
	23903			CTTGTCTTTA		
50	23963			CCTAGGTTCT		
	24023			AGCACTTTGG		
	24083			GCCAGCATGG		
55	24143			GCATGCACCT		
				CONTRACT	CIMALCCCAG	CINCICOIGN

	24203					
		GAAGAATCGC	TTGAAACCAG	GAGGTGTAGG	CTGCAGTGAG	CTGAGATCGC
	24263					
5	24323	TCCAGCCTGG	GCGACAGAAT	GAGACTTTGT	CTCAAAAAA	GAAAAAGATA
	CAACAGGCTA 24383	CCCTTATGTG	CTCACCTTTC	ACTGTTGATT	ACTAGCTATA	AAGTCCTATA
	AAGTTCTTTG	GTCAAGAACC	TTGACAACAC	TAAGAGGGAT	TTGCTTTGAG	AGGTTACTGT
10		TTCATATATA	TACATATACA	TGTATATATG	TATCTATATC	CAGGCTTGGC
		TCAGACTTTC	CAGTGCACTT	GGGAGATGTT	AGGTCAATAT	CAACTTTCCC
	TGGATTCAGA 24623	TTCAACCCCT	TCTGATGTAA	AAAAAAAAA	AAAAAAGAAA	GAAATCCCTT
15		GCACTCAAGT	TTCACCAGGT	GGGGCTTTCC	AAGTTGGGGG	TTCTCCAAGG
	TCATTGGGAT 24743	TGCTTTCACA	TCCATTTGCT	ATGTACCTTC	CCTATGATGG	CTGGGAGTGG
	TCAACATCAA 24803	AACTAGGAAA	GCTACTGCCC	AAGGATGTCC	TTACCTCTAT	TCTGAAATGT
20		TGATTAAAGA	GATTGCCTGT	TCTACCTATC	CACACTCTCG	CTTTCAACTG
	24923				TGAAACGGAG	
25	TCGCCCAGGC 24983	TAGAGTGCAG	TGGCACGATC	TCAGCTCACT	GCAAGCTCTG	CCTCCCGGGT
		CTCCTGCCTC	ACCCTCCCAA	GCAGCTGGGA	CTACAGGCGC	CTGCCACCAT
	GCCCAGCTAA 25103	TTTTTTGTAT	TTTTAGTAGA	GACGGGGTTT	CACCGTGTTA	GCCAGGATGG
30		CTGAACTTGT	GATCCGCCCG	CCTCAGCCTC	CCAAAGTGCT	GGGATTACAG
	GCGTGAGCCA 25223	TCGCACCCGG	CTCAACTGTA	ACTTTCTATA	CTGGTTCATC	TTCCCCTGTA
		AGCTTTTGAA	GTTTTGGCTA	TGGATTATTT	CTCATTTATA	CATTAGATTT
35	CAGATTAGTT 25343	CCAAATTGAT	GCCCACAGCT	TAGGGTCTCT	TCCTAAATTG	TATATTGTAG
		AAGTGGGTGC	CAATAGGGGA	ACTAGTTTAT	ACTTTCATCA	ACTTAGGACC
	CACACTTGTT 25463	GATAAAGAAC	AAAGGTCAAG	AGTTATGACT	ACTGATTCCA	CAACTGATTG
40		GATAACCCCG	TGACCTCTGC	CATCCAGAGT	CTTTCAGGCA	TCTTTGAAGG
	ATGAAGAAAT 25583	GCTATTTTAA	TTTTGGAGGT	TTCTCTATCA	GTGCTTAGGA	TCATGGGAAT
	CTGTGCTGCC 25643	ATGAGGCCAA	AATTAAGTCC	AAAACATCTA	CTGGTTCCAG	GATTAACATG
45		AGGTGGTGCC	CACATGTTCT	GATCCATCCT	GCAAAATAGA	CATGCTGCAC
	TAACAGGAAA 25763	AGTGCAGGCA	GCACTACCAG	TTGGATAACC	TGCAAGATTA	TAGTTTCAAG
50		ATTTCTCACA	AGGCCCTATT	CTGTGACTGA	AACATACAAG	AATCTGCATT
		AGGCAGGGCC	CAGCCAAGGA	GACCATATTC	AGGACAGAAA	TTCAAGACTA
		GGAGTGCTTG	GCAGGGAAGA	CAGAGTCAAG	GACTGCCAAC	TGAGCCAATA
55	CAGCAGGCTT 26003	ACACAGGAAC	CCAGGGCCTA	GCCCTACAAC	AATTATTGGG	TCTATTCACT

	GTAAGTTTTA ATTTCAGGCT CCACTGAAAG AGTAAGCTAA GATTCCTGGC ACTTTCT	GTC
	26063 TCTCTCACAG TTGGCTCAGA AATGAGAACT GGTCAGGCCA GGCATGGTGG CTTACAC	CTG
5	26123	
v	GAATCCCAGC ACTTTGGGAG GCCGAAGTGG GAGGGTCACT TGAGGCCAGG AGTTCAG 26183	GAC
	CAGCTTAGGC AACAAAGTGA GATACCCCCT GACCCCTTCT CTACAAAAAT AAATTTT	AAA.
	AATTAGCCAA ATGTGGTGGT GTATACTTAC AGTCCCAGCT ACTCAGGAGG CTGAGGC	AGG
10	26303	_
	GGGATTGCTT GAGCCCAGGA ATTCAAGGCT GCAGTGAGCT ATGATTTCAC CACTGCA	CTT
	CTGGCTGGGC AACAGAGCGA GACCCTGTCT CAAAGCAAAA AGAAAAAGAA ACTAGAA	CTA
	26423 GCCTAAGTTT GTGGGAGGAG GTCATCATCG TCTTTAGCCG TGAATGGTTA TTATAGA	GGA
15	26483	
	CAGAAATTGA CATTAGCCCA AAAAGCTTGT GGTCTTTGCT GGAACTCTAC TTAATCT 26543	TGA
	GCAAATGTGG ACACCACTCA ATGGGAGAGG AGAGAAGTAA GCTGTTTGAT GTATAGG	GGA
20	AAACTAGAGG CCTGGAACTG AATATGCATC CCATGACAGG GAGAATAGGA GATTCGG	AGT
20	26663 TAAGAAGGAG AGGAGGTCAG TACTGCTGTT CAGAGATTTT TTTTATGTAA CTCTTGA	C 2 2
	26723	
	GCAAAACTAC TTTTGTTCTG TTTGGTAATA TACTTCAAAA CAAACTTCAT ATATTCA 26783	TAAL
25	TGTTCATGTC CTGAAATAAT TAGGTAATGT TTTTTTCTCT ATAG GAA ATG AAT C	CT
	26839 Glu Met Asn P	ro
	85	
	CCT GAT AAC ATC AAG GAT ACA AAA AGT GAC ATC ATA TTC TTT CAG AG 26887	A
30	Pro Asp Asn Ile Lys Asp Thr Lys Ser Asp Ile Ile Phe Phe Gln Ar	g
	90 95 100 AGT GTC CCA GGA CAT GAT AAT AAG ATG CAA TTT GAA TCT TCA TCA TA	.C
	26935 Son Vol. Pro Clu Min Ann Ann Ann Clu Pho Clu Con	
35	Ser Val Pro Gly His Asp Asn Lys Met Gln Phe Glu Ser Ser Ty 105 110 115 12	0
	GAA GGA TAC TTT CTA GCT TGT GAA AAA GAG AGA GAC CTT TTT AAA CT 26983	C
	Glu Gly Tyr Phe Leu Ala Cys Glu Lys Glu Arg Asp Leu Phe Lys Le	u
	125 130 135 ATT TTG AAA AAA GAG GAT GAA TTG GGG GAT AGA TCT ATA ATG TTC AC	ı.Tı
40	27031	
	Ile Leu Lys Lys Glu Asp Glu Leu Gly Asp Arg Ser Ile Met Phe Th	.r
	GTT CAA AAC GAA GAC T AGCTATTAAA ATTTCATGCC GGGCGCAGTG GCTCACG	CCT
	27087 Val Gln Asn Glu Asp	
45	155	
	GTAATCCCAG CCCTTTGGGA GGCTGAGGCG GGCAGATCAC CAGAGGTCAG GTGTTCA 27147	AGA
	CCAGCCTGAC CAACATGGTG AAACCTCATC TCTACTAAAA ATACAAAAAA TTAGCTG	AGT
50	27207 GTAGTGACCC ATGCCCTCAA TCCCAGCTAC TCAAGAGGCT GAGGCAGGAG AATCACT	TGC
	27267	
	ACTCCGGAGG TGGAGGTTGT GGTGAGCCGA GATTGCACCA TTGCGCTCTA GCCTGGG 27327	
	CAACAGCAAA ACTCCATCTC AAAAAATAAA ATAAATAAAT AAACAAATAA AAAATTC 27387	ATA
55	ATGTGAACTG TCTGAATTTT TATGTTTAGA AAGATTATGA GATTATTAGT CTATAAT	TGT

	27447					
	AATGGTGAAA 27507	TAAAATAAAT	ACCAGTCTTG	AAAAACATCA	TTAAGAAATG	AATGAACTTT
5		AACAAACAGA	CTTTCCCTTA	TTTAAGTGAA	TAAAATAAAA	ТААААТАААА
		AAAATTCATA	GTTTGAAAAC	ATTCTACATT	GTTAATTGGC	ATATTAATTA
		AATTATTTT	AAATCTTTTG	GGTTATTAGT	CCTAATGACA	AAAGATATTG
10		TTTCTAATTT	TTAAGAATAT	CGTTAAACCA	TCAATATTTT	TATAAGGAGG
	<del></del> · · - ·	TGACAAATTT	CTGAATTTCC	TCCAAAGTCA	GTATATTTT	AAAATTCAGT
		ATCCAGCAAT	ATATAAAAGG	GATTATATAC	TCTGGCCAAC	TGACATTCAT
15	• • .	CAAAGATGGT	TTAATATCCT	AAAATCAATT	AACATAACAT	ACTATATTAA
		AAACAGTATT	CTCATCTTTT	TTTCTTTTTT	CACAATTCCT	TGGTTACACT
		TAGATGCAGA	AAAAGCATTT	GACAAAATCC	AATTCATAAT	AAAAATTCTC
20		GAGAACATCA	TAAAGGCATC	TATGAAAAAC	CTACAGCTAA	TATCATACTT
		AACTGAATTA	TTTTACCCTA	AGATCAAGAA	TAATGCAAGC	ATGTCAGCTC
25		TATTCAACAT	TGTACTGGAG	GTTCTAGCCA	GAGCAACCAT	ACAATAAATA
25		GCACCCAGAT	TAGAAAGGAA	GTCTTTATTT	GCAGACAACA	TGGTTCTTTA
		CGTCAGGAAT	ACACACACAT	GTTAGAACTA	ATAAGTTCAG	CAAGGTTGCA
30		TCAATATGCA	AAAATACATT	GAAGGCTGGG	CTCAGTGGAG	ATGGCATGTA
		CAGCTACTTG	GGAGGCTGAG	GTAGGAGGAT	CACTTGAGGT	GAGGAGTTTG
		GCAATGTGAT	CTTGCCTGTG	AATAGCCACT	GCACTCGAGC	CTAGGCAACA
35		CCGTCTCCAA	AAAAAAAAT	GGTATATTGG	TATTTCTGTA	TATGAACAAT
	GAATGATCTG 28647	AAAACAAGAA	AATTCCATTC	ACGATGGTAT	TAAAAAAATA	AAATACAAAT
	AAATTTAGCA 28707	AAATAATTAT	AAAACTTGTA	CATCGAAAAT	TTCAAAGCAC	TCTGAGGGAA
40	ATTAAAGATG 28767	ATCTAAATAA	TTGGAGAGAC	ACTCTATGAT	CACTGATTGG	AAAATTCATT
	CAATATTGTT 28827	AAGATAACAA	TTGTCCCCAA	ATTGATGCAT	GCATTCAATT	TAGTCTTCAT
	CAAAATTCCA 28887	GCAGGGTTTT	TGCAGAAATT	GACAAGCTGT	ACCCAAAATG	TATATGGAAA
45	TGAAAAGACC 28947	CAGAAGAGCA	AATAATTTTT	ТАААААСАЛА	GTTGGAAAAC	TTTTACTTCC
	TAATTTTAAA 28994	ACTTACTATA	AACCTAAAGT	TATCAAGACC	ATTTAGT	
50	(16) INFORM	ATTON TOD O	TO TR NO 15			
	(i)SEQ	ATION FOR S UENCE CHARA	CTERISTICS:			
	(	A) LENGTH: 10 B) TYPE: amin	o acid	IS .		
55		D) TOPOLOGY: LECULE TYPE				

(v) FRAGMENT TYPE:N-terminal fragment
(xi) SEQUENCE DESCRIPTION:SEQ ID NO:15:

Tyr Phe Gly Lys Leu Glu Ser Lys Leu Ser 1

## Claims

10

A genomic DNA, which encodes a polypeptide capable of inducing the production of interferon-γ by immunocompetent cells; said polypeptide possessing an amino acid sequence given in SEQ ID NO:1 (where the symbol "Xaa" means "isoleucine" or "threonine") or one of functional equivalents thereof;

15

20

25

30

35

5

## SEQ ID NO: 1:

Tyr 1	Phe	Gly	Lys	Leu 5	Glu	Ser	Lys	Leu	Ser 10	Val	Ile	Arg	Asn		Asn
Asp	Gln	Val	Leu 20	Phe	Ile	Asp	Gln	Gly 25	Asn	Arg	Pro	Leu		Glu	Asp
Met	Thr	Asp 35	Ser	Asp	Cys	Arg	Asp 40	Asn	Ala	Pro	Arg		Ile	Phe	Ile
	50					55					60	Ala			
Ser 65	Val	Lys	Cys	Glu	Lys 70	Ile	Ser	Xaa	Leu	Ser 75	Cys	Glu	Asn	Lys	Ile 80
				85					90					95	Lys
			100					105					110	Asn	
		112					120					125	Ala		
	130					135					140		Asp	Glu	Leu
Gly 145	Asp	Arg	Ser	Ile	Met 150	Phe	Thr	Val	Gln	Asn 155.	Glu	Asp			
	Asp Met Ile Ser 65 Ile Ser Met Lys	Asp Gln Met Thr Ile Ser 50 Ser Val 65 Ile Ser Ser Asp Met Gln Lys Glu 130 Gly Asp	Asp Gln Val  Met Thr Asp 35 Ile Ser Met 50 Ser Val Lys 65 Ile Ser Phe Ser Asp Ile  Met Gln Phe 115 Lys Glu Arg 130 Gly Asp Arg	Asp Gln Val Leu 20  Met Thr Asp Ser 35  Ile Ser Met Tyr 50  Ser Val Lys Cys 65  Ile Ser Phe Lys  Ser Asp Ile Ile 100  Met Gln Phe Glu 115  Lys Glu Arg Asp 130  Gly Asp Arg Ser	Asp Gln Val Leu Phe 20  Met Thr Asp Ser Asp 35  Ile Ser Met Tyr Lys 50  Ser Val Lys Cys Glu 65  Ile Ser Phe Lys Glu 85  Ser Asp Ile Ile Phe 100  Met Gln Phe Glu Ser 115  Lys Glu Arg Asp Leu 130  Gly Asp Arg Ser Ile	Asp Gln Val Leu Phe Ile 20  Met Thr Asp Ser Asp Cys 35  Ile Ser Met Tyr Lys Asp 50  Ser Val Lys Cys Glu Lys 65 70  Ile Ser Phe Lys Glu Met 85  Ser Asp Ile Ile Phe Phe 100  Met Gln Phe Glu Ser Ser 115  Lys Glu Arg Asp Leu Phe 130  Gly Asp Arg Ser Ile Met	Asp Gln Val Leu Phe Ile Asp 20  Met Thr Asp Ser Asp Cys Arg 35  Ile Ser Met Tyr Lys Asp Ser 50  Ser Val Lys Cys Glu Lys Ile 65  Ile Ser Phe Lys Glu Met Asn 85  Ser Asp Ile Ile Phe Phe Gln 100  Met Gln Phe Glu Ser Ser Ser 115  Lys Glu Arg Asp Leu Phe Lys 130  Gly Asp Arg Ser Ile Met Phe	Asp Gln Val Leu Phe Ile Asp Gln 20  Met Thr Asp Ser Asp Cys Arg Asp 35  Ile Ser Met Tyr Lys Asp Ser Gln 50  Ser Val Lys Cys Glu Lys Ile Ser 65  To Ile Ser Phe Lys Glu Met Asn Pro 85  Ser Asp Ile Ile Phe Phe Gln Arg 100  Met Gln Phe Glu Ser Ser Ser Tyr 115  Lys Glu Arg Asp Leu Phe Lys Leu 130  Gly Asp Arg Ser Ile Met Phe Thr	Asp Gln Val Leu Phe Ile Asp Gln Gly 20 25  Met Thr Asp Ser Asp Cys Arg Asp Asn 35 40  Ile Ser Met Tyr Lys Asp Ser Gln Pro 50 55  Ser Val Lys Cys Glu Lys Ile Ser Xaa 65 70  Ile Ser Phe Lys Glu Met Asn Pro Pro 85  Ser Asp Ile Ile Phe Phe Gln Arg Ser 100 105  Met Gln Phe Glu Ser Ser Ser Tyr Glu 115 120  Lys Glu Arg Asp Leu Phe Lys Leu Ile 130 135  Gly Asp Arg Ser Ile Met Phe Thr Val	Asp Gln Val Leu Phe Ile Asp Gln Gly Asn 20	Asp Gln Val Leu Phe Ile Asp Gln Gly Asn Arg 20	Asp Gln Val Leu Phe Ile Asp Gln Gly Asn Arg Pro 20	Asp Gln Val Leu Phe Ile Asp Gln Gly Asn Arg Pro Leu 20	Asp Gln Val Leu Phe Ile Asp Gln Gly Asn Arg Pro Leu Phe 20	Asp Gln Val Leu Phe Ile Asp Gln Gly Asn Arg Pro Leu Phe Glu 20

 The genomic DNA of claim 1, which comprises two or more exons; each of the exons possessing a part of nucleotide sequence given in SEQ ID NO:2;

## SEQ ID NO: 2:

GCCTGGACAG TCAGCAAGGA ATTG	TCTCCC AGTGCATTTT	GCCCTCCTGG CTGCCAACTC	60
TGGCTGCTAA AGCGGCTGCC ACCT	GCTGCA GTCTACACAG	CTTCGGGAAG AGGAAGGAA	120
CCTCAGACCT TCCAGATCGC TTCC	TCTCGC AACAAACTAT	TTGTCGCAGG AATAAAG	177
ATG GCT GCT GAA CCA GTA GA	A GAC AAT TGC ATC	AAC TTT GTG GCA ATG	225

50

45

	Met	Ala -35	Ala	Glu	Pro	Val	Glu -30	Asp	Asn	Cys	Ile	Asn -25	Phe	Val	.Ala	Met	
	AAA	TTT	ATT	GAC	AAT	ACG	CTT	TAC	TTT	ATA	GCT	GAA	GAT	GAT	GAA	AAC	273
5	Lys	Phe	Ile	Asp	Asn	Thr	Leu	Tyr	Phe	Ile	Ala	Glu	Asp	Asp	Glu	Asn	
Ū	-20					-15					-10		_	-		-5	
		GAA															321
	Leu	Glu	Ser	Asp	Tyr 1	Phe	Gly	ГЛ2	Leu 5	Glu	Ser	Lys	Leu	Ser 10	Val	Ile	
		TAA															369
10		Asn	15					20					25				
	CTA	TTT	GAA	GAT	ATG	ACT	GAT	TCT	GAC	TGT	AGA	GAT	AAT	GCA	CCC	CGG	417
	Leu	Phe 30	Glu	Asp	Met	Thr	Asp 35	Ser	Asp	Cys	Arg	Asp 40	Asn	Ala	Pro	Arg	
	ACC	ATA	TTT	ATT	ATA	AGT	ATG	TAT	AAA	GAT	AGC	CAG	CCT	AGA	GGT	ATG	465
15	Thr 45	Ile	Phe	Ile	Ile	Ser 50	Met	Tyr	Lys	Asp	Ser 55	Gln	Pro	Arg	Gly	Met 60	
	GCT	GTA	ACT	ATC	TCT	GTG	AAG	TGT	GAG	AAA	ATT	TCA	AYT	CTC	TCC	TGT	513
•	Ala	Val	Thr	Ile	Ser 65	Val	Lys	Cys	Glu	Lys 70	Ile	Ser	Xaa	Leu	Ser 75	Cys	
	GAG	AAC	AAA	ATT	ATT	TCC	TTT	AAG	GAA	ATG	AAT	CCT	CCT	GAT	AAC	ATC	561
20		Asn		80					85					90			
	AAG	GAT	ACA	AAA	AGT	GAC	ATC	ATA	TTC	TTT	CAG	AGA	AGT	GTC	CCA	GGA	609
		Asp	95					100				_	105			_	
25		GAT															657
	His	Asp 110	Asn	Lys	Met	Gln	Phe 115	Glu	Ser	Ser	Ser	Tyr 120	Glu	Gly	Tyr	Phe	
		GCT															705
	125	Ala				130	_				135				_	140	
30	GAG	GAT	GAA	TTG	GGG	GAT	AGA	TCT	ATA	ATG	TTC	ACT	GTT	CAA	AAC	GAA	753
		Asp			145		_			150					155	Glu	
	GAC	TAGO	TATT	'AA A	TTTA	CATG	C CG	GGCC	CAGI	GGC	TCAC	:GCC	TGTA	ATCC	CA		806
	Asp																
	GCCC	TTTG	GG A	GGCT	GAGG	C GG	GCAG	ATCA	CCA	GAGG	TCA	GGTG	TTCA	AG A	CCAG	CCTGA	866
35	CCAA	CATG	GT G	AAAC	CTCA	T CT	CTAC	TAAA	raa .	'ACTA	AAA	ATTA	GCTG	AG I	'GT'AG	TGACG	926
	CATG		CA A	TUUU	AGCT	A CT	CAAG	AGGC	TGA	GGCA	GGA	GAAT	CACT	TG C	ACTO	CGGAG	986
	GTAG	AGGT	TG T	GG1G	AGCC	G AG	ATTG	LACC	ATT	GCGC	TCT	AGCC	TGGG	CA A	CAAC	AGCAA	1046
					MAT'A	ж АА	'I'AAA	TAAA	TAA	ACAA	ATA	AAAA	ATTC	AT A	ATGT	GAAAA	
	AAAA	AAAA	MAN A	AAA													1120

3. The genomic DNA of claim 1, which comprises two exons with respective nucleotide sequences given in SEQ ID NOs:3 and 4;

## SEQ ID NO: 3:

AA AAC CTG GAA TCA GAT TAC TTT GGC AAG CTT GAA TCT AAA TTA TCA Glu Asn Leu Glu Ser Asp Tyr Phe Gly Lys Leu Glu Ser Lys Leu Ser

55

50

40

5		Val CGG	CCT	Arg CTA	Asn TTT	Leu 15 GAA	AAT Asn GAT Asp	Asp ATG	Gln ACT	Val GAT	Leu 20 TCT	Phe GAC	Ile TGT	Asp AGA	Gln G	GGA Gly 25	10 AAT Asn	95 135
10			SE	Q ID	NO:	4:												
		Asp 40	Asn	Ala	Pro	Arg	ACC Thr 45	Ile	Phe	Ile	Ile	Ser 50	Met	туг	Lys	Asp	Ser 55	47
15		CAG Gln	CCT Pro	AGA Arg	GGT Gly	ATG Met 60	GCT Ala	GTA Val	ACT Thr	ATC Ile	TCT Ser 65	GTG Val	AAG Lys	TGT Cys	GAG Glu	Lys	ATT Ile	95
20		TCA Ser	ACT Thr	CTC Leu	TCC Ser 80	TGT	GAG Glu	AAC Asn	AAA Lys	ATT Ile 85.	ATT	TCC Ser	TTT Phe	AAG Lys		70		134
	4.		genom 5 and		A of cla	aim 1,	which	compr	ises tv	vo exo	ns with	respe	ective (	nucleo	tide se	quenc	es giver	ı in SEQ ID
25			SE	Q ID	NO:	5:												
		GAA	TAAA	G AT Me	G GC t Al -3	a Al	T GA. a Gl	A CC	A GT	A GAA L Glu	ı Asp	AAT Asr	TGC Cys	ATC Ile	AAC Asn -25	Phe	GTG Val	50
30		GCA Ala	ATG Met	AAA Lys -20	TTT	ATT	GAC Asp	AAT Asn	ACG Thr -15	CTT	TAC	TTT Phe	ATA Ile	G Ala -10	-23			87
35				_	NO:													
			Glu		GAT Asp													12
40	5.				A of cla and 6.		which	compr	ises a	ddition	al two	exons	with r	espect	ive nu	cleotid	e seque	nces given
	6.	The g	genom	ic DNA	of cla	im 1, v	hich c	ompris	ses an	exon v	vith a p	art of a	a nucle	otide s	equer	ice giv	en in SE	Q ID NO:7;
45																		
50																		
55																		

SEQ ID NO: 7:

	GAA	ATG	AAT	CCT	CCT	GAT	AAC	ATC	AAG	GAT	ACA	444	λCT	GAC	ATC	A TO A	4.0
5	Glu	Met	Asn	Pro	Pro	Asp	Asp	Tle	Lvs	Asn	Thr	Lve	Ser	Acr	Ile	VIV.	48
3	60					90					95					100	
	TTC	TTT	CAG	AGA	AGT	GTC	CCA	GGA	CAT	GAT	AAT	AAG	ΔΨC	CAA	ттт	CAA	96
	Phe	Phe	Gln	Arg	Ser	Val	Pro	Glv	His	Asp	Asp	Lvs	Met	Gla	Phe	Clu	90
				_	105			1		110		<b>D</b> 13		OIII	115	GIU	
	TCT	TCA	TCA	TAC	GAA	GGA	TAC	ጥጥጥ	CTA	GCT	тст	CAA	444	GAG	AGA	CAC	144
10	Ser	Ser	Ser	Tyr	Glu	Glv	Tvr	Phe	Leu	Ala	Cvs	Glu	Lve	Glu	Arg	ACC.	144
				120					125					130	_	-	
	CTT	TTT	AAA	CTC	ATT	TTG	AAA	AAA	GAG	GAT	GAA	TTG	GGG	CAT	AGA	ጥርጥ	192
	Leu	Phe	Lys	Leu	Ile	Leu	Lys	Lvs	Glu	Asp	Glu	Leu	Glv	ASD	Arg	Ser	174
			135					140					145		_		
10	ATA	ATG	TTC	ACT	GTT	CAA	AAC	GAA	GAC	TAGO	TAT	TAAA	A ጥጥባ	ירא י	ጥርርርር	GGCGC	246
15	Ile	Met	Phe	Thr	Val	Gln	Asn	Glu	Asp							Judeuc	240
		150					155		_								
	AGTO	GCTC	CAC	GCCTC	TAAT	'C CC	AGCC	CTTT	GGG	AGGC	TGA	GGCG	GGCA	GA '	TCACC	CAGAGG	306
	TUAL	90101	TIC 1	AAGAC	CAGC	CTC	ACCA	АСАТ	່ GG1	'GAAA	CCT	CATC	ፈጥጋጥ	CT .		ጥ ለ ር ላ ለ	366
	AAAA	VITAG	CT (	<i>3</i> AGT(	TAGT	'G AC	CCAT	rgccc	TCA	ATCC	CAG	CTAC	TCAA	CA (	SCCTC	ACCCA	176
20	GGAG	BAATC	IAC 1	TTGC	CTCC	G GA	GGTG	GAGG	TTC	TGGT	'GAG	CCGA	CATT	icc :	ТАЭЭА	TOCOC	196
	TCTA	GCCT	GG (	SCAAC	CAACA	G CA	AAAC	CTCCA	TCT	CAAA	AAA	TAAA	ልጥልል	AT I	מתממ	44744	546
	ATAA	AAAA	ALT. C	JATAA	<b>TGTG</b>	A AC	TGTC	TGAA	ጥጥጥ	ጥጥልጥ	CTT	TAGA	AAGA	ጥጥ	ነ ጥር ነ ር	יות איותית אי	606
	TAGT	CTAT	'AA 'I	UTGTA	ATGG	T GA	AATA	TAAA	AAA	TACC	AGT	CTTG	4444	AC I	ነ ጥር አጥ	א ה א חסי	666
	AATG	AATG	AA (	JTTTC	CACAA	A AG	CAAA	CAAA	CAG	ACTT	TCC	CTTA	ጥጥጥ ል	AC C	የር ል ል ጥ	ית א א א י	726
25	AAAA	TAWA	AT	<b>YAAA</b> T	AATG	$\mathbf{T}$	'AAAA	TTAA	CAT	AGTT	TGA	AAAC	ልጥጥር	ጥል (	ን ውጥጥ ረ	יתית א חיתי	706
25	TGGC	ATAT	"I'A A	TATT	ACTT	A AT	ATAA	TATT	TTT	TAAA	TCT	TTTG	GGTT	AT T	PAGTO	CTAAT	846
	GACA	AAAG	AT A	ATTGA	TATT	T GA	ACTT	TCTA	ATT	TTTA	AGA	ATAT	CGTT	AA A	CCAT	CAATA	906
	TTTT	TATA	AG G	AGGC	CACT	T CA	CTTG	ACAA	ATT	TCTG	AAT	TTCC	TCCA	AA C	STCAG	TATAT	966
	CDVC	ለለለለ ጥር እር	את ת	AGTT	TGAT	C CT	GAAT	CCAG	CAA	TATA	TAA	AAGG	GATT	AT A	TACT	CTGGC	1026
	ልሮልጥ	A C TO A C	ער ד	ש גאמיי.	CTAG	G AA	TGCA	AAGA	TGG	TTTA	ATA	TCCT	AAAA	TC A	ATTA	ACATA	1086
30	ጥርርጥ	TOTA	ጥል ር	י ארט אי געט אי	$\pi$ C $\lambda$ $\pi$	T AT		ACAG	TAT	TCTC	ATC	TTTT	TTTC	TT 1	TTTTC	ACAAT	1146
	ТААТ	4444	את ע	ハしょん	Y Y CAL	m Ch	MATA	GATG	CAG	AAAA.	AGC	ATTT	GACA	AA A	TCCA	ATTCA	1206
	CTAA	ጥልጥር	и та	CTCA	ACC A	עם הא	MMGM	CMCA	ATC	ATAA.	AGG	CATC	TATG	AA A	AACC	TACAG	1266
	AAGC	A TICT	U U	CIIA	TCCA	I GA.	MAMA	CTGA	ATT	AT''1''1''	TAC	CCTA.	AGAT	CA A	GAAT	AATGC	1326
	CCAT	ACAA	СЛ С И АТ	ΔΨΔΔ	ABBT	אא מ	YCCC YCIA	TTCA	ACA	TIGI	ACT	GGAG	GTTC	TA G	CCAG	AGCAA	1386
35	AACA	ᡎᢙᢑᡎ	TC ጥ	ጥጥልጥ	CCAC.	ת את.	NGGC.	MC N C	AGA	TTAG	AAA	GGAA	GTCT	TT A	TTTG	CAGAC	1446
00	TCAG	CAAG	ים אים די	יניטי	CMMC	ጥ መጨ የ እአነ	M A M C	1 CAG	GAA	TACA	CAC	ACAT	GTTA	GA A	CTAA	TAAGT TCAGT	1506
	GGAG	ATGG	САТ	OCAG OATO	CTTO	C CTC	CCC N	CCMA	COMP	RAAA	ATA	CATT	SAAG	GC T	GGGC	TCAGT ACTTG	1566
	AGGT	GAGG	AG T	TTCA	GGCT.	<b>υ πν</b>	CTCC	AAMC	CTT	JOUAN TOMBO	366	TGAG	TAG	GA G	GATC	CACTC	1626
	GAGC	CTAG	GC A	ACAA	AGTG:	7 CV		CTCT	CCA		300	TGTG	AATA	30 0	ACTG	ATTTC	1686
	TGTA'	TATG	AA C	AATG	A A TG	ית פילו מיחרי	rcaa	DACA	DCA.	*****	BAA BCC	AAAT	GTA.	ra T	TGGT	AAAAA	1746
40	AATA	AAAT	AC A	AATA	ስልተው፤ ልልጥጥ፣	ר אכו	מממר מממר	MMCM MMN N	TOTAL	4441".	rcc Nom	ATTU	ACGA'	rg g	TATT.	TCAAA	1806
	GCAC'	TCTG	AG G	GAAA	TTAA	A GA'	CAAA.	הגהה ממידים	מתמ	ኒ ለለለለ ኒ ጥጥር/	3 N C	TGTAC		JA A	AATT	ACTGA	1800
	TTGG	AAAA:	TT C	ATTC	AATA	r TG	PTAA!	ATA	D T A A	Z T T C (	יירט יירט	CCAAI	401C	IN T	CATC	CATTC	1926
	AATT	TAGT	CT T	CATC	AAAA'	r TC	CAGC	AGGG	ጥጥጥ	יייים ו	AGA	$\mathbf{c}$	29 C 2 :	או פ	TOTAL	CCCAA	7 A Q Q
	AATG'	TATA	rg g	AAAT	GAAA	A GAG	CCA	GAAG	AGC	AAATI	10A 4	ውም ተመሰው ተመሰው የተመሰው የ የመመር መመር መመር መመር መመር መመር መመር መመር መመር መመር	371 U 71 1 A A A 1	10 C	TOTAL	TTGGA	2106
45	AAAC	rttt/	AC T	TCCT	AATT	TAT 1	AAAC	TTAC	TATE	774A	 ፈጥ	ያ ያ ር ውሳ	ኒ ሌሌሌ/	77 C	AAAG'	TTTAG	2100
	Т										-+11	nag I I	. A I C	יא ט	ACCA:	TTAG	
																	2167.

- 7. The genomic DNA of claim 3, which comprises additional one exon with a part of a nucleotide sequence given in SEQ ID NO:7.
  - 8. The genomic DNA of claim 5, which comprises additional one exon with a part of a nucleotide sequence given in SEQ ID NO:7.
- 9. The genomic DNA of claim 1, which comprises two introns with respective nucleotide sequences given in SEQ ID NOs:8 and 9;

## SEQ ID NO: 8:

	GTATTTTTT		ACATAGAAAT	GACTAGCTAC	TTCTTCCCAT	TCTGTTTTAC	60
5	TGCTTACATT	GTTCCGTGCT		CTCAGATGAA	AAGTCACAGG	AGTGACAATA	120
	ATTTCACTTA	CAGGAAACTT		CCACGTTTTT	TAGTTGGGGT	AAAAAATTGG	180
	ATACAATAAG	ACATTGCTAG	GGGTCATGCC	TCTCTGAGCC	TGCCTTTGAA	TCACCAATCC	240
	CTTTATTGTG	ATTGCATTAA	CTGTTTAAAA	CCTCTATAGT	TGGATGCTTA	ATCCCTGCTT	300
	GTTACAGCTG	AAAATGCTGA		GGTGTGGTGG	CATCTATCTG	TAATCCTAGC	360
10	TACTTGGGAG	GCTCAAGCAG		TGAGGCCAGG	ACTTTGAGGC	TGTAGTACAC	420
10	TGTGATCGTA	CCTGTGAATA	GCCACTGCAC	TCCAGCCTGG	GTGATATACA	GACCTTGTCT	480
	СТААААТТАА		***************************************	TTAGGAAAGG	AAATTGATCA	AGTCTACTGT	540
	GCCTTCCAAA		CAAATATCAA	AGTTAGGCTG	AGTTGAAGCA	GTGAATGTGC	600
	ATTCTTTAAA	AATACTGAAT	ACTTACCTTA	ACATATATTT	TAAATATTTT	ATTTAGCATT	660
	TAAAAGTTAA	AAACAATCTT	TTAGAATTCA	TATCTTTAAA	АТАСТСАААА	AAGTTGCAGC	720
15	GTGTGTGTTG	TAATACACAT	TAAACTGTGG	GGTTGTTTGT	TTGTTTGAGA	TGCAGTTTCA	780
	CTCTGTCACC	CAGGCTGAAG	TGCAGTGCAG	TGCAGTGGTG	TGATCTCGGC	TCACTACAAC	840
	CTCCACCTCC	CACGTTCAAG	CGATTCTCAT	GCCTCAGTCT	CCCGAGTAGG	TGGGATTACA	900
	GGCATGCACC	ACTTACACCC	GGCTAATTTT	TGTATTTTTA	GTAGAGCTGG	GGTTTCACCA	960
	TGTTGGCCAG	GCTGGTCTCA	AACCCCTAAC	CTCAAGTGAT	CTGCCTGCCT	CAGCCTCCCA	1020
	AACAAACAAA	CAACCCCACA	GTTTAATATG	TGTTACAACA	CACATGCTGC	AACTTTTATG	1080
20	AGTATTTTAA	TGATATAGAT	TATAAAAGGT	TGTTTTTAAC	TTTTAAATGC	TGGGATTACA	1140
	GGCATGAGCC	ACTGTGCCAG	GCCTGAACTG	TGTTTTTAAA	AATGTCTGAC		1200
	TAGTCTCCTG	CAGACTGGCC	AAGTCTCAAA	GTGGGAACAG	GTGTATTAAG	GACTATCCTT	1260
	TGGTTAAATT	TCCGCAAATG	TTCCTGTGCA	AGAATTCTTC	TAACTAGAGT	TCTCATTTAT	1320
	TATATTTATT	TCAG					1334
25							

## SEQ ID NO: 9:

GTAAGACTGA GCCTTACTTT GTTTTCAATC ATGTTAATAT AATCAATA	TA ATTAGAAATA 60
TAACATTATT TCTAATGTTA ATATAAGTAA TGTAATTAGA AAACTCAA	AT ATCCTCAGAC 120
CAACCTTTTG TCTAGAACAG AAATAACAAG AAGCAGAGAA CCATTAAA	GT GAATACTTAC 180
TAAAAATTAT CAAACTCTTT ACCTATTGTG ATAATGATGG TTTTTCTC	AG CCTGTCACAG 240
GGGAAGAGGA GATACAACAC TTGTTTTATG ACCTGCATCT CCTGAACA	AT CAGTCTTTAT 300
ACAAATAATA ATGTAGAATA CATATGTGAG TTATACATTT AAGAATAA	CA TGTGACTTTC 360
CAGAATGAGT TCTGCTATGA AGAATGAAGC TAATTATCCT TCTATATT	TC TACACCTTTG 420
TAAATTATGA TAATATTTTA ATCCCTAGTT GTTTTGTTGC TGATCCTT	AG CCTAAGTCTT 480

AGACACAGC TTCAGCTTCC AGTTGATGTA TGTTATTTTT AATGTTAATC TAATTGAATA AAAGTTATGA GATCAGCTGT AAAAGTAATG CTATAATTAT CTTCAAGCCA GGTATAAAGT 600 ATTTCTGGCC TCTACTTTTT CTCTATTATT CTCCATTATT ATTCTCTATT ATTTTCTCT 5 ATTTCCTCCA TTATTGTTAG ATAAACCACA ATTAACTATA GCTACAGACT GAGCCAGTAA GAGTAGCCAG GGATGCTTAC AAATTGGCAA TGCTTCAGAG GAGAATTCCA TGTCATGAAG ACTCTTTTTG AGTGGAGATT TGCCAATAAA TATCCGCTTT CATGCCCACC CAGTCCCCAC TGAAAGACAG TTAGGATATG ACCTTAGTGA AGGTACCAAG GGGCAACTTG GTAGGGAGAA 900 AAAAGCCACT CTAAAATATA ATCCAAGTAA GAACAGTGCA TATGCAACAG ATACAGCCCC 960 CAGACAAATC CCTCAGCTAT CTCCCTCCAA CCAGAGTGCC ACCCCTTCAG GTGACAATTT 1020 10 GGAGTCCCCA TTCTAGACCT GACAGGCAGC TTAGTTATCA AAATAGCATA AGAGGCCTGG 1080 GATGGAAGGG TAGGGTGGAA AGGGTTAAGC ATGCTGTTAC TGAACAACAT AATTAGAAGG 1140 GAAGGAGATG GCCAAGCTCA AGCTATGTGG GATAGAGGAA AACTCAGCTG CAGAGGCAGA 1200 TTCAGAAACT GGGATAAGTC CGAACCTACA GGTGGATTCT TGTTGAGGGA GACTGGTGAA 1260 AATGITAAGA AGATGGAAAT AATGCTTGGC ACTTAGTAGG AACTGGGCAA ATCCATATIT 1320 GGGGGAGCCT GAAGTTTATT CAATTTTGAT GGCCCTTTTA AATAAAAAGA ATGTGGCTGG 1380 15 GCGTGGTGGC TCACACCTGT AATCCCAGCA CTTTGGGAGG CCGAGGGGGG CGGATCACCT 1440 GAAGTCAGGA GTTCAAGACC AGCCTGACCA ACATGGAGAA ACCCCATCTC TACTAAAAAT 1500 ACAAAATTAC CTGGGCGTGG TGGCATATGC CTGTAATCCC AGCTACTCGG GAGGCTGAGG 1560 CAGGAGAATC TTTTGAACCC GGGAGGCAGA GGTTGCGATG AGCCTAGATC GTGCCATTGC 1620 TTAACCAAAG GCATTAGCTT AATAATTTAA TACTGTTTTT AAGTAGGGCG GGGGGTGGCT 1740 20 GGAAGAGATC TGTGTAAATG AGGGAATCTG ACATTTAAGC TTCATCAGCA TCATAGCAAA 1800 TCTGCTTCTG GAAGGAACTC AATAAATATT AGTTGGAGGG GGGGAGAGAG TGAGGGGTGG 1860 ACTAGGACCA GTTTTAGCCC TTGTCTTTAA TCCCTTTTCC TGCCACTAAT AAGGATCTTA 1920 GCAGTGGTTA TAAAAGTGGC CTAGGTTCTA GATAATAAGA TACAACAGGC CAGGCACAGT 1980 GGCTCATGCC TATAATCCCA GCACTTTGGG AGGGCAAGGC CAGTGTCTCA CTTGAGATCA 2040 25 GGAGTTCAAG ACCAGCCTGG CCAGCATGGC GATACTCTGT CTCTACTAAA AAAAATACAA 2100 AAATTAGCCA GGCATGGTGG CATGCACCTG TAATCCCAGC TACTCGTGAG CCTGAGGCAG 2160 AAGAATCGCT TGAAACCAGG AGGTGTAGGC TGCAGTGAGC TGAGATCGCA CCACTGCACT 2220 CCAGCCTGGG CGACAGAATG AGACTTTGTC TCAAAAAAAG AAAAAGATAC AACAGGCTAC 2280 CCTTATGTGC TCACCTTTCA CTGTTGATTA CTAGCTATAA AGTCCTATAA AGTTCTTTGG 2340 TCAAGAACCT TGACAACACT AAGAGGGATT TGCTTTGAGA GGTTACTGTC AGAGTCTGTT 2400 30 TCATATATAT ACATATACAT GTATATATGT ATCTATATCC AGGCTTGGCC AGGGTTCCCT 2460 CAGACTTTCC AGTGCACTTG GGAGATGTTA GGTCAATATC AACTTTCCCT GGATTCAGAT 2520 TCAACCCCTT CTGATGTAAA AAAAAAAAA AAAAGAAAG AAATCCCTTT CCCCTTGGAG 2580 CACTCAAGTT TCACCAGGTG GGGCTTTCCA AGTTGGGGGT TCTCCAAGGT CATTGGGATT 2640 GCTTTCACAT CCATTTGCTA TGTACCTTCC CTATGATGGC TGGGAGTGGT CAACATCAAA 2700 ACTAGGAAAG CTACTGCCCA AGGATGTCCT TACCTCTATT CTGAAATGTG CAATAAGTGT 2760 35 GATTAAAGAG ATTGCCTGTT CTACCTATCC ACACTCTCGC TTTCAACTGT AACTTTCTTT 2820 TTTTCTTTTT TTCTTTTTT CTTTTTTTT GAAACGAGT CTCGCTCTGT CGCCCAGGCT 2880 AGAGTGCAGT GGCACGATCT CAGCTCACTG CAAGCTCTGC CTCCCGGGTT CACGCCATTC 2940 TCCTGCCTCA CCCTCCCAAG CAGCTGGGAC TACAGGCGCC TGCCACCATG CCCAGCTAAT 3000 TTTTTGTATT TTTAGTAGAG ACGGGGTTTC ACCGTGTTAG CCAGGATGGT CTCGATCTCC 3060 TGAACTTGTG ATCCGCCCGC CTCAGCCTCC CAAAGTGCTG GGATTACAGG CGTCAGCCAT 3120 40 CGCACCCGGC TCAACTGTAA CTTTCTATAC TGGTTCATCT TCCCCTGTAA TGTTACTAGA 3180 GCTTTTGAAG TTTTGGCTAT GGATTATTTC TCATTTATAC ATTAGATTTC AGATTAGTTC 3240 CAAATTGATG CCCACAGCTT AGGGTCTCTT CCTAAATTGT ATATTGTAGA CAGCTGCAGA 3300 ACTGGGTGCC AATAGGGGAA CTAGTTTATA CTTTCATCAA CTTAGGACCC ACACTTGTTG 3360 ATAAAGAACA AAGGTCAAGA GTTATGACTA CTGATTCCAC AACTGATTGA GAAGTTGGAG 3420 45 ATAACCCCGT GACCTCTGCC ATCCAGAGTC TTTCAGGCAT CTTTGAAGGA TGAAGAAATG 3480 CTATTTTAAT TTTGGAGGTT TCTCTATCAG TGCTTAGGAT CATGGGAATC TGTGCTGCCA 3540 TGAGGCCAAA ATTAAGTCCA AAACATCTAC TGGTTCCAGG ATTAACATGG AAGAACCTTA 3600 GGTGGTGCCC ACATGTTCTG ATCCATCCTG CAAAATAGAC ATGCTGCACT AACAGGAAAA 3660 GTGCAGGCAG CACTACCAGT TGGATAACCT GCAAGATTAT AGTTTCAAGT AATCTAACCA 3720 TTTCTCACAA GGCCCTATTC TGTGACTGAA ACATACAAGA ATCTGCATTT GGCCTTCTAA 3780 50 GGCAGGGCCC AGCCAAGGAG ACCATATTCA GGACAGAAAT TCAAGACTAC TATGGAACTG 3840 GAGTGCTTGG CAGGGAAGAC AGAGTCAAGG ACTGCCAACT GAGCCAATAC AGCAGGCTTA 3900

5	TTTCAGGCTC TGGCTCAGAA CTTTGGGAGG ACAAAGTGAG TGTGGTGGTG AGCCCAGGAA ACAGAGCGAG TGGGAGGAGG ATTAGCCCAA CACCACTCAA CTGGAACTGA GGAGGTCAGT TTTGTTCTGT	CACTGAAAGA ATGAGAACTG CCGAAGTGGG ATACCCCCTG TATACTTACA TTCAAGGCTG ACCCTGTCTC TCATCATCGT AAAGCTTGTG TGGGAGAGGA ATATGCATCC ACTGCTGTTC TTGGTAATAT	GTAAGCTAAG GTCAGGCCAG AGGGTCACTT ACCCCTTCTC GTCCCAGCTA CAGTGAGCTA AAAGCAAAAA CTTTAGCCGT GTCTTTGCTG GAGAAGTAAG CATGACAGGG AGAGATTTTT ACTTCAAAAC	ATTCCTGGCA GCATGGTGGC GAGGCCAGGA TACAAAATA CTCAGGAGGC TGATTTCACC GAAAAAGAAA GAATGGTTAT GAACTCTACT CTGTTTGATG AGAATAGGAG TTTATGTAAC AAACTTCATA	CTTTCTGTCT TTACACCTGG GTTCAGGACC AATTTTAAAA TGAGGCAGGG ACTGCACTTC CTAGAACTAG TATAGAGGAC TAATCTTGAG TATAGGGGAA ATTCGGAGTT TCTTGAGAAG	TGGCTGGCA CCTAAGTTTG AGAAATTGAC CAAATGTGGA AACTAGAGGC AAGAAGGAGA	4020 4080 4140 4200 4260 4320 4380 4440 4500 4560 4620 4680 4740
15		AGGTAATGTT			INTICAANTI	GIICKIGICC	4773.

10. The genomic DNA of claim 1, which comprises three introns with respective nucleotide sequences of SEQID NOs: 10 to 12 as introns;

## SEQ ID NO: 10:

GTAAGAAATA	TCATTCCTCT	TTATTTGGAA	AGTCAGCCAT	GGCAATTAGA	GGTAAATAAG	60
CTAGAAAGCA	ATTGAGAGGA	АТАТАААССА		CTACGATGAG		120
CAACATAAGA	AATATAAGCA	AAGTCAGAGT	AGAATTTTTT	TCTTTTATCA	GATATGGGAG	180
AGTATCACTT	TAGAGGAGAG	GTTCTCAAAC	TTTTTGCTCT	CATGTTCCCT	TTACACTAAG	240
CACATCACAT	GTTAGCATAA	GTAACATTTT	TAATTAAAA	TAACTATGTA	CTTTTTTAAC	300
AACAAAAAA	AGCATAAAGA	GTGACACTTT	TTTATTTTA	CAAGTGTTTT	AACTGGTTTA	360
ATAGAAGCCA	TATAGATCTG	CTGGATTCTC	ATCTGCTTTG	CATTCAGACT	ACTGCAATAT	420
TGCACAGAAT	GCAGCCTCTG	GTAAACTCTG	TTGTACACTC	ATGAGAGAAT	GGGTGAAAAA	480
GACAAATTAC	GTCTTAGAAT	TATTAGAAAT	AGCTTTCACT	TTAGGAACTC	CCTGAGAATT	540
GCTGCTTTAG	AGTGGTAAGA	TAAATAAGCT	TCTCTTTAAA	CGGAATCTCA	AGACAGAATC	600
AGTTACATTA	AAAGCAAACA	AAAAATTTGC	CCATGGTTAG	TCATCTTGTG	AAATCTGCCA	660
CACCTTTGGA	CTGGGCTACA	ATTGGATAAT	ATAGCATTCC	CCGAGATAAT	TTTCTCTCAC	720
AATTAAGGAA	AGGGCTGAAT	AAATATCTCT	GTTTGAAGTT	GAATAACAAA	AATTAGGACC	780
CCCTAAATTT	TAGGGCTCCT	GAAATTCGTC	TTTTTGCCTA	TATTCAGCTA	CTTTACGTTC	840
ТАТТАААТСТ	TCTTTCAGGC	CAGGTGCACT	AGCTCATGCC	TAGAATCTCA	GGCAGGCCTG	900
AGCCCAGGAA	TTTGAGACCA	GCCAGGGCAA	CACAGTCTCT	ACAAAAAAAT	AAAAAATTAC	960
CTGGGTGTGT	TGGTGCATGC	CTGTAGAACT	ACTCAGGATG	CTGAGGACTG	CTTGAGCCCA	1020
GGATAGCCAA	ATCTGTGGTG	AGTTCAGCCA	CTAAACAGAG	CGAGACTTTC	TCAAAAAAAC	1080
AAACAAAAA	ACAAACAAAC	TTCCTTCAAA	ATAACTTTTT	ATCTGCAATG	TTTTCCTATT	1140
GCCTGTGAGA	TTAAATTTAC	TCTTTTACCT	GATTTCCAAA	GCCCTCCATA	ATCTAATCCG	1200
ACTTTACCTT	GTGTTCACTG	CAAAATAGCA	GGACTGTTCC	ACTACAATCC	AAAAATCACA	1260
GGTTGGGTGC	AGTGGCTCAC	TCCTGTAATC	CCAACACTTT	GGAAGGCCAA	GGCAGGTGGA	1320
TTGCTTCAGC		AAGACCAGCC		GGCAAAAACC	CTGTCTCTCC	1380
	AAAATTAGCC	AGATGTGGTA	GTATGTGCCT	GTAGTCCCAA	CTACTCAAAA	1440
GGCTAAGGCA	AGAGGATCAC	TTGAGCCCAG	GAGGTCAAGG	CTACAGTGAG	CCATGTTTAC	1500
TGTGTCACTG	CACTCCAGCC	TGGGTGATAG	AGCAAGACCA	TGTCTCAAAA	AAAAAAAAA	1560
GAAAAGAAAA	0		CAGTTCACCC	CCACCACAAC	ATTGTTTTGA	1620
	AATGCTGGTC	CATTGCCTTC	TCTATCTATT	CAAATCTTTA	AGCATTCTTT	1680
GAGATTCAAC	TCAATTCTCC	TTTTCAAACT	AGGCCATTTA	AACTACATCA	GTTCCATTTT	1740

```
GATTTTCTTG CTTTGAGTCT ACAGACTCAA AAACAAAAAC TTAAAAAACTT ATTTTTTAAG 1800
     TTTTCTGCTA CTCTCACTTC TTCAACACTC ACATACACGC ATTCATAATA AGATGGCAGA 1860
     ATGTTCAAGG ATAAAATGAT TTATAGAACT GAAAAGTTAG GTTTTGATCT TGTTGCTGTC 1920
     AAGATGACTA CCTACCTGAT CTCAGGTAAT TAATTATGTA GCATGCTCCC TCATTTCATC 1980
     CCATACCTAT TCAACAGGAT TGGAATTCCA CAGCAAGGAT AAACATAATC ATAGTTGCTT 2040
     TTCAAGTTCA AGGCATTTTA ACTTTTAATC TAGTAGTATG TTTGTTGTTGT 2100
     TTGAGATGGA GCCCTGCTGT GTCACCCAGG CTGGAGTGCA GTGGCACGAA CTCGGCTCAC 2160
     TGCAACCTCT GCCTCATGGG TTCAATCAGT TATTCTGCCT CAGTGTCCCA AGTAGCTGGG 2220
     ACTACAAGGC ACATGCCACC ATGCCTGGCT AATTTTTGTA TTTTTAGTAG AAACAGGGCT 2280
10
     TCACCATGTT GGCCAGGCTG GTCTCGAACT CCTGACCTCA AGTGATCCAG CCGCCTCGGC 2340
     CTCCCAAAGT GCTGGGATTA CAGGCATAAG CCACCGTGCC CAGCCTAATA GTATGTTTTT 2400
     AAACTCTTAG TGGCTTAACA ATGCTGGTTG TATAATAAAT ATGCCATAAA TATTTACTGT 2460
     CTTAGAATTA TGAAGAAGTG GTTACTAGGC CGTTTGCCAC ATATCAATGG TTCTCTCCTT 2520
     ACAGCTTTAA TTAGAGTCTA GAATTGCAGG TTGGTAGAGC TGGAACAGAC CTTAAAGATT 2580
     GACTAGCCAA CTTCCTTGTC CAAATGAGGG AACTGAGACC CTTAAAATTA AGTGACTTGC 2640
15
     CCCAGACAAA ACTGGAACTC ATGTGTCCTA ATTTCCATCA TGAAATTCTA CCATTCACTA 2700
     GCCTCTGGCT AGTTGTCAAA GTATTGCATA ACTAAATTTT TATGTCTGTT TTAAAGAACA 2760
     AATTGTCACT GCTTACTCCT GGGAGGGTCT TTCTGAGGTG GTTTATAACT CTTAAAAAAA 2820
     AAAAAGTCAG TAGTCTGAGA ATTTTAGACG AAATAGTCAA AGCATTTTTA TCCAATGGAT 2880
     CTATAATTT CATAGATTAG AGTTAAATCA AAGAAACACG GATGAGAAAG GAAGAGGAAA 2940
20
     ATTGAGGAGA GGAGGAATGG GGATGAGAAC ACACTACTTG TAATCAGTCA TAGATGTACT 3000
     GAGAACTAAC AAGAAGAATT GTAAGAAAAT AAGAATGAAG AATTCAAAAT CAACACATGA 3060
     AATAAAAAGA AACTACTAGG GAAAAATGGA GAAGACATTA GAAAAATTAT TCTATTTTTA 3120
     AAATTCTGTT TTCAGGCTTC CCTCCTGTTC TTCCTCCTTC TCATTGGTTT TCAGGTGGAG 3180
     GGAAAGTTTA AGATGGAAAA AATATATATA TTCTACACAT CCCTTTCTAC GCTGTTGTCA 3240
     TGGCAACAAG GTTTATCATA GCAAACTTTT ATTCATACAA CATTTATTGA GTTCTTACTG 3300
     TGTGGTAAGC TCTTTCCAGG TGTTGAAAAT TCAGGGGAAA AAAGACAACT CATTGTCTTA 3360
     AAACTCAGAT GAAAGCTGAA CAGACCTATT TTTAATCAAA GTAATCTCAA TTTAGGGTAG 3420
     TAAGAGCTAT TTAAGAAGCA TGAACAGGTG TGAAGGAGGT AGGACTCTGA GGAGAGAATA 3480
     GTTAGCTAGG AATGAAAGAG CAGAGAAGTT TTCCTAGAGG AACTATTAAA GCTGGGAGTT 3540
     ACGGGATGAA AGATGAGGCA GGGTTTGCAG GCAAAAAAAA AAAAAAGGCA GGGGAAGGGG 3600
30
     AAGTTCTGGC CTGGCAGAGA GAATAACTGT GGCAACAATG GAGGAGAGTC TGGAAGCAAG 3660
     AAAACCAAGT AGAAGAGTAT TAAAATAGAA GATGCCAGGG GTAATGAGGG CTTGATTTAA 3720
     AACAGTGCTG TTGGAGATGG AGAGGAGATA CCAAATTCTG GAGACATTTC TGAGTTAGAA 3780
     CCTACAGTAT TTATCAGACA AGGGAAAGAT TAGACAAAGG AGTTAAGAAT GACTCCCAGG 3840
     TTTCAGTTTG GGGCAGGTAA CTAGGACATG TTTTGAAAAG TAATGTATTG GATCTCTTAC 3900
     CATTGGAACT ATGTATGTGG AGCCAAATTA AAATTTGTAC ATGTATATAA CTCTCCCCC 3960
35
     ACCACCAGTA ACTACTTCCC TAACTCTCTA CTTTGTAGCC AGACTTCCTA AAAGAATAGT 4020
     TTGTAGTCAC TGTCTTTACT TTTCCCCTCC CATTCTGTCC TAGATATTTG TCCACCTACC 4080
    ATCTGCTGCC TCCACTTTAC CCAAACTGTT CTACGGTTGC CCAAAACTTC CTAATTGCCA 4140
     AATTCAATGA ACAAGTTTAA GCTTATATGT AAATTAGGAG CTCTACAGTT TGATTTCGAG 4200
    CAGCCCCTCC TGAAACCCTT TCTCTTTCGA CTTCTGTGAC ACATCTCAGA TTTACAAAAC 4260
    TGAACTAATT ATTTTACACT TGAGCTGTAT TTTCGTTCTT CTTTCTTGAT GAATGAGGTA 4320
    ACCACTCAAC AAATTGCCCA AGCCAAAAAC TACGAAGTCA TCCTCAGTTC CTCCTTCTTC 4380
    TGTTTGACCC ACAACAGATC AGCTGAGAAA TCCCGCTGTT TAGTATCTCT TGAATTCATT 4440
    ACCTTAATTT ATAGCCTCAT CAACTCTTAA TTGTTAAAAT TACTTCAGTA GTTGTTGTCT 4500
    GACCTCTGTC CAATCTTGTT CAATCAGGTC CATTCTTTTG TTCTTGGTGG TGGTGGTGGT 4560
    GTTGACAGAG TTTCGCTTTT GCTGCCCAGG CTGAAGTGCA GTGGAGCACT TCACTGCAAC 4620
45
    CACAGCCTCC TGGGTTTAAG CAGTTCACCC TCCCGAGTAG CTGGGACTAC AGGTATGTGC 4680
    CACCACACCC AGCTAATTTT GTGTTTCAG TAGAGACAGG GTTTCACCAT GTTGGTCAGG 4740
    CTGGTCTCAA ACTCCTGACC TCAAGCAATC CACCCACCTC AGCCTCCCAA AGTGCTGGGA 4800
    TTACAGGCAT GAGCCACTGC ACACGGACCA GATCCATTGT TTATGTTGCT TCTAGAGTGA 4860
    GTTTTTAAAA CACAAATTTG ACCATATCTT TCTCCAATTT AAGTCAGTAT TTTTTTTTTC 4920
    AGGAAAAAAC AGTTCAAACT CTTTAGTCTG CTTACACAAG GCCTTTGTAG TCTGACTCTT 4980
    CTTTCCAAGC TTTCATCAAA GTATACTGCA AGTTACATTT TATGTGAATT GAATTAGGCA 5040
    ACGGTATAAA AATTATAGTT TATATGGGCA AAATGGAAAT AATGTTAACT CTTCCAAATA 5100
    GTTTATCTAG AATGACATAA TTTCAAAGCT GTCAGGTCAA ATGAGTTATA AACTGTTAAC 5160
```

```
ACTATTGCCA CATGCAAGTG TCTCTTATAC TTGGTAGAAT TATCTGCTTC CATGTCATTA 5220
       TTATGTAAAT TAGACTTTAA ATAACTCAGA AGTTCTTCAG ACATACAGGT TATTATTGTG 5280
       CTTTTTAAAC ATAATTTTAA ATAATTTTAT ATATGATAAT GTTATCCAAG TGCTAAGGGA 5340
5
       TGTATTGTTA CTGCTGCA AAAAAAAAAA AAAAAAAAC TCCAAATAAA TATGTTGAAA 5400
       CCAAGTTTAT ATGCAAGAAA ACAATATTAA AAAGGCCAAA GTACCACCAT AATAGGCTGT 5460
      GTGGAGACGG CAGGCTACAA AACACTAGTA ATAATGCTGA GAAAGTTGAA AAAAGAAAGA 5520
       AAGCAACAAT ATGCTTTGGT TGTTGTAGGT TTATGTACTC CAAGAATATC TCCTCTCAAA 5580
       CTTTTACGTT TTTTCCAAAG AAAAGTTAAC TTTGGCTGGG CGCAGTGGCT CTTGCCTGTA 5640
       GTCCCAGCCT TTGGGAGGCC AAGGCGGGCA GATCACCTGA GGTCAGGAGT TTGAGACCAG 5700
10
       CCTGACCAAA AATGGAGAAA CCCGCCCCCC TCACTACTAA AAGAATACAA AATTAGGCCG 5760
       GGCACAGTGG CTTACCCCTG TGATCCCAGC ACTTTGGGAG GCCGAAGCAG GAAGATCACC 5820
       TGAGGTCAGG AGTTCGAGAC CAGCCATGGA GAAACCCGTC TCTACTAAAA ATACAAAATT 5880
       AGCCGGGCGT GGTGGTGCAT GACTGTAATC CCAGCTACTC AGGAGGCTAA GGCAGAGAAT 5940
       CACTTGAACC CAGGCAGTGG AGGTTGCAGT GAGCCGAGAT CGTGCCATTG CACTCCAGCC 6000
15
       ACTATGTGAG ATCTTTAGAA ATGCATTCTT TCTGTAAAAT GTGACTACAT TTGCCTTATT 6120 TATGCTAAAA ATGTTGAGGC CTCAAACAAC CCATATTTTC TCGGTCTCCC CGCTGCCTAG 6180
      CCTTTGTTCA CATTGCTTCT TCTTGGTGGA AGCTCTTCCT CTGGCCTTGA AAATGCCTGC 6240
      TTCTCTTTCA AGGTAGCACA GTCATCACTT TCTGTGGTAA CCTTCTCCAG CACCATCAAA 6300
      CAGAAAGAAT GAATCTCTTG TAAATTCAGC TCTTACGTCA TTCATTACAT TATTTTGTAA 6360
20
      CTCTTTATAG ATTCTTCTCT CCCACTAGAC TCTGAGTCAC TGGAGAGTAG GAGCCAACTC 6420
      TCATTCATGT GTGGTTTGGT CAGCTACTGG CCACATTCCT GATGCATAGT TAATGCTCAA 6480
      ACCTTAACTG GTGAATCAGC TCAAATATTG TCCTTCTCTA AATCCATTCA CTCATTGACT 6540
      AACTATGTAC TCAAAATAGT AAACACCAGT AATTTAATCC AATTCCTGCC CATACTGCTT 6600
      GGTACATTTC AGGTGAATTA GTTTGATAAA TATGTGTGTA TTACATAATA TTAAAGTATG 6660
      TACAGAAGAT CATGCTAATC ATAATTCACA ACTGATAACT AATCAAACAT AAATGCTCTC 6720
25
      AGGTTAACAA ATGTCTGCCT TCTCAGTTAA TGCAGTCATT AACAAACACC TTCTGATGCT 6780
      GATAATAGGG CCTTGTTCAG CAATGAAGCC ATAAAGGTGA ATAAAGAACA TGCCCTCGTG 6840
      GAGCTCACAG CCTAGTCATT ATTGTTCTGA TTTTTAATAT TAATGTTGGT TTGGGTTTTG 6900
      GTGAAAAATG TTTAGACTTA TCTTAGTGAT CTTTTCATCC TTTGCTATAT TATTTTTCTC 6960
      TAAGAGTCTT CCTTATCCCC TCCTTTAAAA AACTAGGTGA TAATTCTAAA TTGTAAATTT 7020
30
      AAATATTATA AATAGCTTAT AAAATTTAAT ATTTATAATA TTTAAATGTT TGATAAATAT 7080
      TTTAAATGTG TTGGCCAGGC ATGGTGGCTG ACACCTATAA TCCCAGAACT TTGAGAGGCC 7200
      AAGTCAGGCA AACCATTTGA GCTCAGGAGT TTGAGACCAC CCTGGGCAAC GTGGTGAAAC 7260
      CCTGTCTCTA CCAAACATAT GAAAACTTAT CTGGGTGTGG TGGCACGCAT CTGTGGTCCC 7320
      35
      GGATGTTGCA GTGAGCTGAG ATCGTGCCAC TGCACTCCAA CCTGGGTGAC AGAGTGAGAC 7440
      TCCATCTCAA AAAAAAAAA TGTTATCTAA ATAAGATAAA TTTAATAACT GTTCGCACTT 7500
      AGATGAGCAT AAGGAACTAA ACCTAGATAA AACTATCAAA TAAGGCCTGG GTACAGTGAC 7560
      TCATGCCTGT AATCTCAAGC ACTTTGGGAG GCCAAAATTA TACAAAGTTA GTTGTATAAC 7620
      ACCAACTAAC AACTATTTTG GGGTTAGCTT AATTCAGATT AATTTTTTT AAACTGAGTT 7680 TTAAATTCCT GCTTACTCTA CCATACATGC TAGGCCTCAT ATTATGCTAG AAAAATTTTG 7740
40
      AGCACAGATT TATGAATACT CTCCTGCATA CCATTTAATT TTTAAACAAA TTTTAATGCA 7800
      GTATATATGT GCCTTTTTAC CAACACATTA AATAATAAGA TCTACTGTGA GGACTAAATT 7860
      TCTGTAATTT CAAAGTAGTA ATGAGTTTAA ACCATGTCTC AAGATCTCTG CAATAACTGT 7920
      AGCACAACAG AAAATAGGTA TTTCTATTAA TGACAGAGTC ACAAGTACTA CTAATAATAC 7980
      TGTGGTTTGT TTCCTGCAAC TAATCATGGG AGGAATGCTA AATTTCAGAG GTTGGTGAAA 8040
45
      ATACATGTGT ATTTTTTCC CCATCCAAGT TCACAGATTT CTCACACTGA GAACTCCTAT 8100
      TCCATAACAA AATTCTGGAA GCCTGCACAC CGTATTGGAA GAAGGGCAGA AAGGAAAAGC 8160
AAATGGAAGG ATTTAAATTT TTTTCAAATC CTGTATCCCT TGATTTTACA GCAAGATTGT 8220
ATTTATGTAT TACTTGTGTT AAAAATATAG TATAATCGAG ACTCCAGATC AAAAATCACC 8280
      GCAGCTCAGG GAGAAAGAGG GCCACCAAAT GCCAGAGCCC TTCAGCCTTC TCCCACCCTG 8340
      CCTGTACCCT CAGATGGAAG CACTTTTTTA TCATTGTTTC ACCTTTAGCA TTTTGACAAT 8400
50
      GAAGTCACAA ACCTTCAGCC TCTCACCCAT AGGAACCCAC TGGTTGTAAG AGAAGGATGA 8460
      AGCCAGTCCT TCCTAAAGGG CACGATTAGA TGTGTTTATG GCATCCTCAG GTGAAACTAT 8520
```

```
TACTAGGCCA TTTATCTACC CTTTATAATA TTGTTTAATG AGAAAATGCT TTCTATCTTC 8640
      CAAATATCTG ATGATTTGTA AGAGAACACT TAAACATGGG TATTCATAAG CTGAAACTTC 8700
      TGGCATTTAT TGAATGTCAA GATTGTTCAT CAGTATACTA GGTGATTAAC TGACCACTGA 8760
5
      ACTTGAAGGT AGTATAAAGT AGTAGTAAAA GGTACAATCA TTGTCTCTTA ACAGATGGCT 8820
      CTTTGCTTTC ATTAG
           SEQ ID NO: 11:
10
      GTAAGGCTAA TGCCATAGAA CAAATACCAG GTTCAGATAA ATCTATTCAA TTAGAAAAGA
      TGTTGTGAGG TGAACTATTA AGTGACTCTT TGTGTCACCA AATTTCACTG TAATATTAAT 120
      GGCTCTTAAA AAAATAGTGG ACCTCTAGAA ATTAACCACA ACATGTCCAA GGTCTCAGCA 180
      CCTTGTCACA CCACGTGTCC TGGCACTTTA ATCAGCAGTA GCTCACTCTC CAGTTGGCAG 240
      TAAGTGCACA TCATGAAAAT CCCAGTTTTC ATGGGAAAAT CCCAGTTTTC ATTGGATTTC 300
15
      CATGGGAAAA ATCCCAGTAC AAAACTGGGT GCATTCAGGA AATACAATTT CCCAAAGCAA 360
      ATTGGCAAAT TATGTAAGAG ATTCTCTAAA TTTAGAGTTC CGTGAATTAC ACCATTTTAT
                                                                         420
      GTAAATATGT TTGACAAGTA AAAATTGATT CTTTTTTTTT TTTTCTGTTG CCCAGGCTGG
                                                                         480
      AGTGCAGTGG CACAATCTCT GCTCACTGCA ACCTCCACCT CCTGGGTTCA AGCAATTCTC
                                                                          540
      CTGCCTCAGC CTTCTGAGTA GCTGGGACTA CAGGTGCATC CCGCCATGCC TGGCTAATTT
                                                                         600
      TTGGGTATTT TTACTAGAGA CAGGGTTTTG GCATGTTGTC CAGGCTGGTC TTGGACTCCT 660
20
      GATCTCAGAT GATCCTCCTG GCTCGGGCTC CCAAAGTGCT GGGATTACAG GCATGAACCA
      CCACACATGG CCTAAAAATT GATTCTTATG ATTAATCTCC TGTGAACAAT TTGGCTTCAT 780
      TTGAAAGTTT GCCTTCATTT GAAACCTTCA TTTAAAAGCC TGAGCAACAA AGTGAGACCC 840
      CATCTCTACA AAAAACTGCA AAATATCCTG TGGACACCTC CTACCTTCTG TGGAGGCTGA 900
      AGCAGGAGGA TCACTTGAGC CTAGGAATTT GAGCCTGCAG TGAGCTATGA TCCCACCCCT 960
25
      ACACTCCAGC CTGCATGACA GTAGACCCTG ACACACACA ACAAAAAAAA ACCTTCATAA 1020
      AAAATTATTA GTTGACTTTT CTTAGGTGAC TTTCCGTTTA AGCAATAAAT TTAAAAGTAA 1080
      AATCTCTAAT TTTAGAAAAT TTATTTTTAG TTACATATTG AAATTTTTAA ACCCTAGGTT 1140
      TAAGTTTTAT GTCTAAATTA CCTGAGAACA CACTAAGTCT GATAAGCTTC ATTTTATGGG 1200
      CCTTTTGGAT GATTATATA TATTCTGATG AAAGCCAAGA CAGACCCTTA AACCATAAAA 1260
      ATAGGAGTTC GAGAAAGAGG AGTAGCAAAA GTAAAAGCTA GAATGAGATT GAATTCTGAG 1320
      TCGAAATACA AAATTTTACA TATTCTGTTT CTCTCTTTTT CCCCCTCTTA G
           SEQ ID NO: 12:
      GTAAAGTAGA AATGAATTTA TTTTTCTTTG CAAACTAAGT ATCTGCTTGA GACACATCTA
35
      TCTCACCATT GTCAGCTGAG GAAAAAAAA AATGGTTCTC ATGCTACCAA TCTGCCTTCA 120
      AAGAAATGTG GACTCAGTAG CACAGCTTTG GAATGAAGAT GATCATAAGA GATACAAAGA 180
      AGAACCTCTA GCAAAAGATG CTTCTCTATG CCTTAAAAAA TTCTCCAGCT CTTAGAATCT 240
      ACAAAATAGA CTTTGCCTGT TTCATTGGTC CTAAGATTAG CATGAAGCCA TGGATTCTGT
                                                                          300
      TGTAGGGGGA GCGTTGCATA GGAAAAAGGG ATTGAAGCAT TAGAATTGTC CAAAATCAGT
40
      AACACCTCCT CTCAGAAATG CTTTGGGAAG AAGCCTGGAA GGTTCCGGGT TGGTGGTGGG
      GTGGGGCAGA AAATTCTGGA AGTAGAGGAG ATAGGAATGG GTGGGGCAAG AAGACCACAT
                                                                         480
      TCAGAGGCCA AAAGCTGAAA GAAACCATGG CATTTATGAT GAATTCAGGG TAATTCAGAA
                                                                         540
      TGGAAGTAGA GTAGGAGTAG GAGACTGGTG AGAGGAGCTA GAGTGATAAA CAGGGTGTAG
                                                                         600
      AGCAAGACGT TCTCTCACCC CAAGATGTGA AATTTGGACT TTATCTTGGA GATAATAGGG
                                                                         660
      TTAATTAAGC ACAATATGTA TTAGCTAGGG TAAAGATTAG TTTGTTGTAA CAAAGACATC
                                                                         720
      CAAAGATACA GTAGCTGAAT AAGATAGAGA ATTTTTCTCT CAAAGAAAGT CTAAGTAGGC
                                                                         780
      AGCTCAGAAG TAGTATGGCT GGAAGCAACC TGATGATATT GGGACCCCCA ACCTTCTTCA
      GTCTTGTACC CATCATCCCC TAGTTGTTGA TCTCACTCAC ATAGTTGAAA ATCATCATAC
      TTCCTGGGTT CATATCCCAG TTATCAAGAA AGGGTCAAGA GAAGTCAGGC TCATTCCTTT
      CAAAGACTCT AATTGGAAGT TAAACACATC AATCCCCCTC ATATTCCATT GACTAGAATT 1020
      TAATCACATG GCCACACCAA GTGCAAGGAA ATCTGGAAAA TATAATCTTT ATTCCAGGTA 1080
50
      GCCATATGAC TCTTTAAAAT TCAGAAATAA TATATTTTTA AAATATCATT CTGGCTTTGG 1140
      TATAAAGAAT TGATGGTGTG GGGTGAGGAG GCCAAAATTA AGGGTTGAGA GCCTATTATT 1200
      TTAGTTATTA CAAGAAATGA TGGTGTCATG AATTAAGGTA GACATAGGGG AGTGCTGATG 1260
```

```
AGGAGCTGTG AATGGATTTT AGAAACACTT GAGAGAATCA ATAGGACATG ATTTAGGGTT 1320
      GGATTTGGAA AGGAGAAGAA AGTAGAAAAG ATGATGCCTA CATTTTTCAC TTAGGCAATT 1380
5
      TGTACCATTC AGTGAAATAG GGAACACAGG AGGAAGAGCA GGTTTTGGTG TATACAAAGA 1440
      GGAGGATGCA TGACGCATTT CGTTTTGGAT CTGAGATGTC TGTGGAACGT CCTAGTGGAG 1500
      ATCTCCACAA ACTCTTCTAC ATGTGGTTCT GAGTTCAGGA CACAGATTTG GGCTGGAGAT 1560
      AGAGATATTG TAGGCTTATA CATAGAAATG GCATTTGAAT CTATAGAGAT AAAAAGACAC 1620
      ATCAGAGGAA ATGTGTAAAG TGAGAGAGGA AAAGCCAAGT ACTGTGCTGG GGGGAATACC 1680
      TACATTTAAA GGATGCAGTA GAAAGAAGCT AATAAACAAC AGAGAGCAGA CTAACCAAAA 1740
10
      GGGGAGAAGA AAAACCAAGA GAATTCCACC GACTCCCAGG AGAGCATTTC AAGATTGAGG 1800
      GGATAGGTGT TGTGTTGAAT TTTGCAGCCT TGAGAATCAA GGGCCAGAAC ACAGCTTTTA 1860
      GATTTAGCAA CAAGGAGTTT GGTGATCTCA GTGAAAGCAG CTTGATGGTG AAATGGAGGC 1920
      AGAGGCAGAT TGCAATGAGT GAAACAGTGA ATGGGAAGTG AAGAAATGAT ACAGATAATT 1980
      CTTGCTAAAA GCTTGGCTGT TAAAAGGAGG AGAGAAACAA GACTAGCTGC AAAGTGAGAT 2040
      TGGGTTGATG GAGCAGTTTT AAATCTCAAA ATAAAGAGCT TTGTGCTTTT TTGATTATGA 2100
15
      AAATAATGTG TTAATTGTAA CTAATTGAGG CAATGAAAAA AGATAATAAT ATGAAAGATA 2160
      AAAATATAAA AACCACCCAG AAATAATGAT AGCTACCATT TTGATACAAT ATTTCTACAC 2220
      TCCTTTCTAT GTATATAC AGACACAGAA ATGCTTATAT TTTTATTAAA AGGGATTGTA 2280
      CTATACCTAA GCTGCTTTTT CTAGTTAGTG ATATATATGG ACATCTCTCC ATGGCAACGA 2340
      GTAATTGCAG TTATATTAAG TTCATGATAT TTCACAATAA GGGCATATCT TTGCCCTTTT 2400
      TATTTAATCA ATTCTTAATT GGTGAATGTT TGTTTCCAGT TTGTTGTTGT TATTAACAAT 2460
20
      GTTCCCATAA GCATTCCTGT ACACCAATGT TCACACATTT GTCTGATTTT TTCTTCAGGA 2520
      TAAAACCCAG GAGGTAGAAT TGCTGGGTTG ATAGAAGAGA AAGGATGATT GCCAAATTAA 2580
      AGCTTCAGTA GAGGGTACAT GCCGAGCACA AATGGGATCA GCCCTAGATA CCAGAAATGG 2640
      CACTTTCTCA TTTCCCCTTG GGACAAAAGG GAGAGAGGCA ATAACTGTGC TGCCAGAGTT 2700
      AAATTTGTAC GTGGAGTAGC AGGAAATCAT TTGCTGAAAA TGAAAACAGA GATGATGTTG 2760
25
      TAGAGCTCCT GAAGAGACCA AAGAAAATTT GAAATTGCGG CTATCAGCTA TGGAAGAGAG 2820
      TGCTGAACTG GAAAACAAAA GAAGTATTGA CAATTGGTAT GCTTGTAATG GCACCGATTT 2880
      GAACGCTTGT GCCATTGTTC ACCAGCAGCA CTCAGCAGCC AAGTTTGGAG TTTTGTAGCA 2940
      GAAAGACAAA TAAGTTAGGG ATTTAATATC CTGGCCAAAT GGTAGACAAA ATGAACTCTG 3C00
      AGATCCAGCT GCACAGGGAA GGAAGGGAAG ACGGGAAGAG GTTAGATAGG AAATACAAGA 3060
      GTCAGGAGAC TGGAAGATGT TGTGATATTT AAGAACACAT AGAGTTGGAC TAAAAGTGTA 3120
30
      AGAAAACTAG AAGGGTAAGA GACCGGTCAG AAAGTAGGCT ATTTGAAGTT AACACTTCAG 3180
      AGGCAGAGTA GTTCTGAATG GTAACAAGAA ATTGAGTGTG CCTTTGAGAG TAGGTTAAAA 3240
      AACAATAGGC AACTTTATTG TAGCTACTTC TGGAACAGAA GATTGTCATT AATAGTTTTA 3300
      GAAAACTAAA ATATATAGCA TACTTATTTG TCAATTAACA AAGAAACTAT GTATTTTTAA 3360
      ATGAGATTTA ATGTTTATTG TAG
                                                                         3383.
```

- 11. The genomic DNA of claim 9, which comprises additional three introns with respective nucleotide sequences given in SEQ ID NOs:10 to 12.
- 40 12. The genomic DNA of claim 1, which has a nucleotide sequence selected from the group consisting of SEQ ID NO: 13 and its complementary sequence;

SEQ ID NO: 13:

AAG ATG GCT GAA CCA GTA GAA GAC AAT TGC ATC AAC TTT GTG GCA
Met Ala Ala Glu Pro Val Glu Asp Asn Cys Ile Asn Phe Val Ala

55

50

35

		35	20		_		
			-30		-25		
	Met Lus Pl	he Tle Ace	AAT ACG CTT	TAC TTT A	TA G GTAAG	G CTAATGCCAT	98
5	-20	ne lie ASp	Asn Thr Leu	Tyr Phe I.			
3		a CCACCMMCA	-15		-10		
	ATTAACTCA	~ CCAGGIICA	G ATAAATCTA	T TCAATTAG	A AAGATGTTG	r GAGGTGAACT	158
	GTGGACCTC	L VCJIJGIGI.	C ACCAAATTT	C ACTGTAATA	TAATGGCTC	TAAAAAAATA	218
	GTCCTGCCA	T HOUNTILWY	C CACAACATG	T CCAAGGTCT	C AGCACCTTG	CACACCACGT	278
	AAATCCCAG		C AGTAGCTCA	C TCTCCAGT	G GCAGTAAGT	G CACATCATGA	338
10	GTACAAAAC	P GGGTCCAMM	A AAATUUCAG	TTTTCATTG	A TTTCCATGG	AAAAATCCCA	398
	AGAGATTCTC	C GGOIGCAII	L AGGAAATAC	A ATTTCCCA	A GCAAATTGG	AAAAATCCCA AAATTATGTA	458
	ACTAAAAAATT	CATTCTTAG	A GTTCCGTGA	A TTACACCAT	T TTATGTAAA	ATGTTTGACA	518
	СТСТССТСАС	TECANCETC	TTTTTTTTC	r GTTGCCCAG	G CTGGAGTGC	GTGGCACAAT	578
	AGTAGCTGGG	0133AA331 S	CARCCOCC	TTCAAGCAA	T TCTCCTGCC1	CAGCCTTCTG	638
	GAGACAGGGT	TTTTCCCAMC	T TOTOCO COCO	TGCCTGGCT	A ATTTTTGGGT	ATTTTACTA	698
15	CCTGGCTCGG	GCTCCCAAA	TOTOCAGGC:	r GGTCTTGGA	C TCCTGATCTC	AGATGATCCT	758
	AATTGATTCT	TATCATTA A	T COCCOCCO	P ACAGGCATG	A ACCACCACAC	ATGCCTAAA	818
	ATTTGAAACC	·	ACCCMCACO	A CAATTIGGC	T TCATTTGAAA	GTTTGCCTTC	878
	TGCAAAATAT	CCTGTGGAC	AGCCIGAGCA	ACAAAGTGA	G ACCCCATCTO	TACAAAAAAC	938
	GAGCCTAGGA	ATTTCACCC	CCICCIACC	TCTGTGGAG	G CTGAAGCAGC	AGGATCACTT	998
20	GACAGTAGAC	CCTGACACAC	CACACACAAA	ATGATCCCA	С СССТАСАСТО С АТААААААТТ	CAGCCTGCAT	1058
	TTTTCTTAGG	TGACTTTCCC	- ACACACAAAA - TTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTT	AAAAACCTT	С АТААААААТТ А СТААААТСТС	ATTAGTTGAC	1118
	AAATTTATTT	TTAGTTACAT	, 1117777777	MANTITAAA	A GTAAAATCTC A GGTTTAAGTT	TAATTTTAGA	1178
	ATTACCTGAG	AACACACTA	GTCTCATAILT	COMMCAMMON	A GGTTTAAGTT A TGGGCCTTTT	TTATGTCTAA	1238
	ATAATATTCT	GATGAAAGCC	, PECICATAMO	COMPARACON	A TGGGCCTTTT T AAAAATAGGA	GGATGATTAT	1298
	GAGGAGTAGC	AAAAGTAAAA		CARMONARCO	T AAAAATAGGA C TGAGTCGAAA	GTTCGAGAAA	1358
25	TACATATTCT	GTTTCTCTCT	, GC17G471G4	CTTAC	C TGAGTCGAAA T GAA GAT GA	TACAAAATTT	1418
		0111010101	11110000	CITAG C	r GAA GAT GA	T G GTAAA	1470
		•		-1	a Glu Asp As	p Glu	
	GTAGAAATGA	ATTTATTTT	CTTTGCAAAC	TAAGTATCT	- CTTCACACAC	A THOM A THOM ON	1520
	CCATIGICAG	CIGAGGAAAA	. AAAAAAATGG	TTCTCATCC	ኮ አርርአአጥርጥርር	COUNTRY A S C S S	1530 1590
30	AIGIGGACIC	AGTAGCACAG	CTTTGGAATG	AAGATGATC	4 TAACACATAC	A A A C A A C A A C	1650
	CICIAGCAAA	AGATGCTTCT	CTATGCCTTA	. ልልልልልጥጥሮጥ/	~ CACCTCTTAC	A A TC T A C A A A	1710
	ATAGACTTIG	CUTGTTTCAT	TGGTCCTAAG	ATTAGCATG	1 ACCCAMCCAM	TO TO THE TOTAL OF	1770
	GGGGAGCGTT	GCATAGGAAA	<b>AAGGGATTGA</b>	AGCATTAGA	1 TTCTCCAAAA	TORONACAC	1830
	CICCICICAG	MANTGCTTTG	GGAAGAAGCC	<b>TGGAAGGTT</b> (	CGGGTTGGTG	CTCCCCTCCC	1890
	GCAGAAAATT	CTGGAAGTAG	AGGAGATAGG	AATGGGTGG	CCAAGAAGAC	CACAMMCACA	1950
35	GGCCAAAAGC	TGAAAGAAAC	CATGGCATTT	<b>ATGATGAAT</b> 1	' CACCCTAATT	CACAAMCCAA	2010
	GINGAGIAGG	AGTAGGAGAC	TGGTGAGAGG	AGCTAGAGTO	2 ATAAACACCC	TOTAL CAROLA	2070
	GACGIICICI	CACCCCAAGA	TGTGAAATTT	GGACTTTATO	TTCCACATA	TACCCOMD A A M	2130
	TANGCACAAT	ATGTATTAGC	TAGGGTAAAG	<b>ልጥጥልርጥጥጥር</b> ባ	י תכתאארהאאר	ACAMCCA AAC	2190
	ATACAGTAGE	TGAATAAGAT	AGAGAATTTT	TCTCTCAAAC	2 AAACTCTAAC	TACCCACCMC	2250
40	AGAAGTAGTA	TGGCTGGAAG	CAACCTGATG	ATATTCCCAC	' CCCC & & CC TOTO	CMMC & CMCMM	2310
	GIACCCATCA	TCCCCTAGTT	GTTGATCTCA	CTCACATAGT	תמשמשת השחים	CAMACOMOCOM	2370
	COGLICATAL	CCCAGTTATC	AAGAAAGGGT	CAAGAGAAGT	, Caccentamm	CCMMMCAAAO	2430
	MOTOTANTIG	GAAGTTAAAC	ACATCAATCC	CCCTCATATI	' ("ርልጥጥርልርጥአ	CAAMMMAAMO	2490
	ACA I GGCCAC	ACCAAGTGCA	AGGAAATCTG	GAAAATATAA		A COM A COCA M	2550
	AIGACICITY	AAAATTCAGA	AATAATATAT	<b>ΥΥΥΥΝΑΑΑΑΥΝ</b>	TCATTCTCCC	THE COURT OF A	2610
45	VOVVIIGNIC	GIGIGGGIG	AGGAGGCCAA	AATTAAGGGT	TCACACCCTA	TO A TO TO TO TO A COM	2670
	INTINCANGA	AATGATGGTG	TCATGAATTA	AGGTAGACAT	AGGGGAGTGC	TO A TICA COA C	2730
	CIGIGANIGG	ATTTTAGAAA	CACTTGAGAG	<b>AATCAATAGG</b>	<b>ልቦልጥርልጥጥጥ</b> አ	CCCMMCCAMM	2790
	IGGANAGGAG	AAGAAAGTAG	AAAAGATGAT	GCCTACATTT	<b>ጥጥሮ ልሮጥጥ ልሮር</b>	CAAMMMCMAC	2850
	CHITCHGIGH	AATAGGGAAC	ACAGGAGGAA	GAGCAGGTTT	ጥርርጥርጥልጥልር	222222242444	2910
50	A I GOA I GACG	CATTTCGTT	TGGATCTGAG	ATGTCTGTGG	AACGTCCTAG	TOCACATORO	2970
	CACAAACTCT	TUTACATGTG	GTTCTGAGTT	CAGGACACAG	ATTTCCCCTC	CACAMACACA	3030
	INTIGINGGE	TTATACATAG	AAATGGCATT	TGAATCTATA	GAGATAAAA	CACACAMCAC	3090
	AUGAAATGTG	TAAAGTGAGA	GAGGAAAAGC	CAAGTACTGT	GCTGGGGGGA	ATACCTACAT	3150

```
TTAAAGGATG CAGTAGAAAG AAGCTAATAA ACAACAGAGA GCAGACTAAC CAAAAGGGGA
                                                                             3210
       GAAGAAAAC CAAGAGAATT CCACCGACTC CCAGGAGAGC ATTTCAAGAT TGAGGGGATA
GGTGTTGTGT TGAATTTTGC AGCCTTGAGA ATCAAGGGCC AGAACACAGC TTTTAGATTT
                                                                             3270
                                                                             3330
       AGCAACAAGG AGTTTGGTGA TCTCAGTGAA AGCAGCTTGA TGGTGAAATG GAGGCAGAGG
5
                                                                             3390
       CAGATTGCAA TGAGTGAAAC AGTGAATGGG AAGTGAAGAA ATGATACAGA TAATTCTTGC
                                                                             3450
       TAAAAGCTTG GCTGTTAAAA GGAGGAGAGA AACAAGACTA GCTGCAAAGT GAGATTGGGT
                                                                             3510
       TGATGGAGCA GTTTTAAATC TCAAAATAAA GAGCTTTGTG CTTTTTTGAT TATGAAAATA ATGTGTTAAT TGTAACTAAT TGAGGCAATG AAAAAAGATA ATAATATGAA AGATAAAAAT
                                                                             3570
       ATAAAAACCA CCCAGAAATA ATGATAGCTA CCATTTTGAT ACAATATTTC TACACTCCTT
                                                                             3630
       TCTATGTATA TATACAGACA CAGAAATGCT TATATTTTTA TTAAAAGGGA TTGTACTATA
10
       CCTAAGCTGC TTTTTCTAGT TAGTGATATA TATGGACATC TCTCCATGGC AACGAGTAAT
       TGCAGTTATA TTAAGTTCAT GATATTTCAC AATAAGGGCA TATCTTTGCC CTTTTTATTT
                                                                             3870
      AATCAATTCT TAATTGGTGA ATGTTTGTTT CCAGTTTGTT GTTGTTATTA ACAATGTTCC
                                                                             3930
      CATAAGCATT CCTGTACACC AATGTTCACA CATTTGTCTG ATTTTTTCTT CAGGATAAAA
                                                                             3990
       CCCAGGAGGT AGAATTGCTG GGTTGATAGA AGAGAAAGGA TGATTGCCAA ATTAAAGCTT
                                                                             4050
15
      CAGTAGAGGG TACATGCCGA GCACAAATGG GATCAGCCCT AGATACCAGA AATGGCACTT
                                                                             4110
      TCTCATTTCC CCTTGGGACA AAAGGGAGAG AGGCAATAAC TGTGCTGCCA GAGTTAAATT
                                                                             4170
       TGTACGTGGA GTAGCAGGAA ATCATTTGCT GAAAATGAAA ACAGAGATGA TGTTGTAGAG
                                                                             4230
      GTCCTGAAGA GAGCAAAGAA AATTTGAAAT TGCGGCTATC AGCTATGGAA GAGAGTGCTG
                                                                             4290
      AACTGGAAAA CAAAAGAAGT ATTGACAATT GGTATGCTTG TAATGGCACC GATTTGAACG
                                                                             4350
      CTTGTGCCAT TGTTCACCAG CAGCACTCAG CAGCCAAGTT TGGAGTTTTG TAGCAGAAAG
20
                                                                             4410
      ACAAATAAGT TAGGGATTTA ATATCCTGGC CAAATGGTAG ACAAAATGAA CTCTGAGATC
                                                                             4470
      CAGCTGCACA GGGAAGGAAG GGAAGACGGG AAGAGGTTAG ATAGGAAATA CAAGAGTCAG
                                                                            4530
      GAGACTGGAA GATGTTGTGA TATTTAAGAA CACATAGAGT TGGAGTAAAA GTGTAAGAAA
                                                                            4590
      ACTAGAAGGG TAAGAGACCG GTCAGAAAGT AGGCTATTTG AAGTTAACAC TTCAGAGGCA
                                                                            4650
      GAGTAGTTCT GAATGGTAAC AAGAAATTGA GTGTGCCTTT GAGAGTAGGT TAAAAAACAA
                                                                            4710
      TAGGCAACTT TATTGTAGCT ACTTCTGGAA CAGAAGATTG TCATTAATAG TTTTAGAAAA
                                                                            4770
      CTAAAATATA TAGCATACTT ATTTGTCAAT TAACAAAGAA ACTATGTATT TTTAAATGAG
                                                                            4830
      ATTTAATGTT TATTGTAG AA AAC CTG GAA TCA GAT TAC TTT GGC AAG CTT
                                                                            4880
                           Glu Asn Leu Glu Ser Asp Tyr Phe Gly Lys Leu
                               -5
      GAA TCT AAA TTA TCA GTC ATA AGA AAT TTG AAT GAC CAA GTT CTC TTC
                                                                            4928
30
      Glu Ser Lys Leu Ser Val Ile Arg Asn Leu Asn Asp Gln Val Leu Phe
                       10
                                           15
      ATT GAC CAA GGA AAT CGG CCT CTA TTT GAA GAT ATG ACT GAT TCT GAC
                                                                            4976
      Ile Asp Gln Gly Asn Arg Pro Leu Phe Glu Asp Met Thr Asp Ser Asp
                  25
                                       30
                                                           35
                  GTATTTTT TTAATTCGCA AACATAGAAA TGACTAGCTA CTTCTTCCCA
      TGT AGA G
35
                                                                            5032
      Cys Arg Asp
              40
      TTCTGTTTTA CTGCTTACAT TGTTCCGTGC TAGTCCCAAT CCTCAGATGA AAAGTCACAG
                                                                            5092
      GAGTGACAAT AATTTCACTT ACAGGAAACT TTATAAGGCA TCCACGTTTT TTAGTTGGGG
                                                                            5152
      TAAAAAATTG GATACAATAA GACATTGCTA GGGGTCATGC CTCTCTGAGC CTGCCTTTGA
                                                                            5212
      ATCACCAATC CCTTTATTGT GATTGCATTA ACTGTTTAAA ACCTCTATAG TTGGATGCTT
40
      AATCCCTGCT TGTTACAGCT GAAAATGCTG ATAGTTTACC AGGTGTGGTG GCATCTATCT
                                                                            5332
      GTAATCCTAG CTACTTGGGA GGCTCAAGCA GGAGGATTGC TTGAGGCCAG GACTTTGAGG
      CTGTAGTACA CTGTGATCGT ACCTGTGAAT AGCCACTGCA CTCCAGCCTG GGTGATATAC
                                                                            5452
      AGACCTTGTC TCTAAAATTA AAAAAAAAA AAAAAAAAC CTTAGGAAAG GAAATTGATC
                                                                            5512
      AAGTCTACTG TGCCTTCCAA AACATGAATT CCAAATATCA AAGTTAGGCT GAGTTGAAGC
45
      AGTGAATGTG CATTCTTTAA AAATACTGAA TACTTACCTT AACATATATT TTAAATATTT
                                                                            5632
      TATTTAGCAT TTAAAAGTTA AAAACAATCT TTTAGAATTC ATATCTTTAA AATACTCAAA
                                                                            5692
      5752
                                                                            5812
      CTCACTACAA CCTCCACCTC CCACGTTCAA GCGATTCTCA TGCCTCAGTC TCCCGAGTAG
                                                                            5872
      GTGGGATTAC AGGCATGCAC CACTTACACC CGGCTAATTT TTGTATTTTT AGTAGAGCTG
50
                                                                            5932
      GGGTTTCACC ATGTTGGCCA GGCTGGTCTC AAACCCCTAA CCTCAAGTGA TCTGCCTGCC
                                                                            5992
      TCAGCCTCCC AAACAAACAA ACAACCCCAC AGTTTAATAT GTGTTACAAC ACACATGCTG
```

5	CAACTTTAT GAGTATTTA ATGATATAGA TTATAAAAGG TTGTTTTAA CTTTTAAATG CTGGGATTAC AGGCATGAGC CACTGTGCCA GGCCTGAACT GTGTTTTTAA AAATGTCTGA CCAGCTGTAC ATAGTCTCCT GCAGACTGGC CAAGTCTCAA AGTGGGAACA GGTGTATTAA	6112 6172 6232
	GGACTATCCT TTGGTTAAAT TTCCGCAAAT GTTCCTGTGC AAGAATTCTT CTAACTAGAG	6292
	TTCTCATTTA TTATATTTAT TTCAG AT AAT GCA CCC CGG ACC ATA TTT ATT	6343
	Asp Asn Ala Pro Arg Thr Ile Phe Ile 40 45	
10	ATA AGT ATG TAT AAA GAT AGC CAG CCT AGA GGT ATG GCT GTA ACT ATC	6391
10	Ile Ser Met Tyr Lys Asp Ser Gln Pro Arg Gly Met Ala Val Thr Ile 50 55 60	
	TCT GTG AAG TGT GAG AAA ATT TCA ACT CTC TCC TGT GAG AAC AAA ATT	6439
	Ser Val Lys Cys Glu Lys Ile Ser Thr Leu Ser Cys Glu Asn Lys Ile	
	65 70 75 80 ATT TCC TTT AAG GTAAG ACTGAGCCTT ACTTTGTTTT CAATCATGTT AATATAATCA	6496
15	The Ser Phe Lys	0470
	ATATAATTAG AAATATAACA TTATTTCTAA TGTTAATATA AGTAATGTAA TTAGAAAACT	6556
	CAAATATCCT CAGACCAACC TTTTGTCTAG AACAGAAATA ACAAGAACCA GAGAACCATT	6616
	AAAGTGAATA CTTACTAAAA ATTATCAAAC TCTTTACCTA TTGTGATAAT GATGGTTTTT	6676
	CTGAGCCTGT CACAGGGGAA GAGGAGATAC AACACTTGTT TTATGACCTG CATCTCCTGA	6736
20	ACAATCAGTC TTTATACAAA TAATAATGTA GAATACATAT GTGAGTTATA CATTTAAGAA	6796
	TAACATGTGA CTTTCCAGAA TGAGTTCTGC TATGAAGAAT GAAGCTAATT ATCCTTCTAT	6856
	ATTTCTACAC CTTTGTAAAT TATGATAATA TTTTAATCCC TAGTTGTTTT GTTGCTGATC CTTAGCCTAA GTCTTAGACA CAAGCTTCAG CTTCCAGTTG ATGTATGTTA TTTTTAATGT	6916 6976
	TAATCTAATT CAATAAAAGT TATGAGATCA GCTGTAAAAG TAATGCTATA ATTATCTTCA	7036
	AGCCAGGTAT AAAGTATTC TGGCCTCTAC TTTTTCTCTA TTATTCTCCA TTATTATTCT	7096
25	CTATTATTTT TCTCTATTTC CTCCATTATT GTTAGATAAA CCACAATTAA CTATAGCTAC	7156
20	AGACTGAGCC AGTAAGAGTA GCCAGGGATG CTTACAAATT GGCAATGCTT CAGAGGAGAA	7216
	TTCCATGTCA TGAAGACTCT TTTTGAGTGG AGATTTGCCA ATAAATATCC GCTTTCATGC	7276
	CCACCCAGTO CCCACTGAAA GACAGTTAGG ATATGACCTT AGTGAAGGTA CCAAGGGGCA	7336
	ACTTGGTAGG GAGAAAAAAG CCACTCTAAA ATATAATCCA AGTAAGAACA GTGCATATGC	7396
	AACAGATACA GCCCCCAGAC AAATCCCTCA GCTATCTCCC TCCAACCAGA GTGCCACCCC	7456
30	TTCAGGTGAC AATTTGGAGT CCCCATTCTA GACCTGACAG GCAGCTTAGT TATCAAAATA GCATAAGAGG CCTGGGATGG AAGGGTAGGG TGGAAAGGGT TAAGCATGCT GTTACTGAAC	7516 7576
	AACATAATTA GAAGGGAAGG AGATGGCCAA GCTCAAGCTA TGTGGGATAG AGGAAAACTC	7636
	AGCTGCAGAG GCAGATTCAG AAACTGGGAT AAGTCCGAAC CTACAGGTGG ATTCTTGTTG	7696
	AGGGAGACTO GTGAAAATGT TAAGAAGATG GAAATAATGC TTGGCACTTA GTAGGAACTG	7756
	GGCAAATCCA TATTTGGGGG AGCCTGAAGT TTATTCAATT TTGATGGCCC TTTTAAATAA	7816
35	AAAGAATGTG GCTGGGCGTG GTGGCTCACA CCTGTAATCC CAGCACTTTG GGAGGCCGAG	7876
	GGGGGCGGAT CACCTGAAGT CAGGAGTTCA AGACCAGCCT GACCAACATG GAGAAACCCC	7936
	ATCTCTACTA AAAATACAAA ATTAGCTGGG CGTGGTGGCA TATGCCTGTA ATCCCAGCTA	7996
	CTCGGGAGGC TGAGGCAGGA GAATCTTTTG AACCCGGGAG GCAGAGGTTG CGATGAGCCT	8056
	AGATCGTGCC ATTGCACTCC AGCCTGGGCA ACAAGAGCAA AACTCGGTCT CAAAAAAAA AAAAAAAAAG TGAAATTAAC CAAAGGCATT AGCTTAATAA TTTAATACTG TTTTTAAGTA	8116 8176
40	GGGCGGGGG TGGCTGGAAG AGATCTGTGT AAATGAGGGA ATCTGACATT TAAGCTTCAT	8236
	CAGCATCATA GCAAATCTGC TTCTGGAAGG AACTCAATAA ATATTAGTTG GAGGGGGGGA	8296
	GAGAGTGAGG GGTGGACTAG GACCAGTTTT AGCCCTTGTC TTTAATCCCT TTTCCTGCCA	8356
	CTAATAAGGA TCTTAGCAGT GGTTATAAAA GTGGCCTAGG TTCTAGATAA TAAGATACAA	8416
	CAGGCCAGGC ACAGTGGCTC ATGCCTATAA TCCCAGCACT TTGGGAGGGC AAGGCGAGTG	8476
45	TCTCACTTGA GATCAGGAGT TCAAGACCAG CCTGGCCAGC ATGGCGATAC TCTGTCTCTA	8536
43	CTAAAAAAA TACAAAAATT AGCCAGGCAT GGTGGCATGC ACCTGTAATC CCAGCTACTC	8596
	GTGAGCCTGA GGCAGAAGAA TCGCTTGAAA CCAGGAGGTG TAGGCTGCAG TGAGCTGAGA	8656
	TCGCACCACT GCACTCCAGC CTGGGCGACA GAATGAGACT TTGTCTCAAA AAAAGAAAAA	8716
	GATACAACAG GCTACCCTTA TGTGCTCACC TTTCACTGTT GATTACTAGC TATAAAGTCC TATAAAGTTC TTTGGTCAAG AACCTTGACA ACACTAAGAG GGATTTGCTT TGAGAGGTTA	8776
	CTGTCAGAGT CTGTTTCATA TATATACATA TACATGTATA TATGTATCTA TATCCAGGCT	8836 8896
50	TGGCCAGGGT TCCCTCAGAC TTTCCAGTGC ACTTGGGAGA TGTTAGGTCA ATATCAACTT	8956
	TCCCTGGATT CAGATTCAAC CCCTTCTGAT GTAAAAAAAA AAAAAAAAA GAAAGAAATC	9016

	CC	TTTCC	CCT	TGG	AGCAG	CTC .	AAGT:	PTCAC	CC AC	GGTG	GGGC1	ר דד י	CAAC	STTG	GGGG	STTCTCC	9076
	AAG	SCTCF	TTG	GGA	rtgcz	rtt (	CACA	rccai	T TO	GCTAT	rgtac	CTI	ירררייו	באתם	ስጥርር	CTCCCA	0126
	GT	JGTC#	ACA	TCA	AAACI	CAG (	GAAAC	SCTAC	CT GO	CCCAA	AGGAT	CTC	ጉጥጥላ	ጉርጥ	CTA	מ א שחיים	0106
5	ATO	5'1'GC#	AT'A	AGTO	STGAT	ΓTA .	AAGAC	SATTO	C C1	ჁႺႥႥჇ	CTACC	: ጥልባ	ירראר	ጥጋልና	CTCC		0254
	AC	<b>IGTA</b>	CTT	TCTT	ודדדי	TTC '	rttt	TTTCT	ייד ידי	יייייי	րահան	ւ ահահան	ישייהי	200	CCAC	STCTCGC	9316
	TC	CTCC	CCC	AGGC	CTAGA	AGT (	GCAG	CGCA	C G	<b>አጥርጥር</b>	AGCT	CAC	יתיה ביי יתיה ביי	2000	TCTC	CCTCCC	9310
	GGC	STTCA	CGC	CATT	CTCC	TG (	ירייר ג	יייייייייייייייייייייייייייייייייייייי	יר כנ	2000	מכריי	CCC	. y C.w.y	CAC	1010	CTGCCA	9376
	CCZ	TGCC	CAG	CTA	ሊጥጥጥባ	י יחייניי	2010. 2010	ית חייתיים זיים חייתיים	C T	CACA		Con		CAG	GCGC	GCCAGG	9436
	ATC	GTCT	CGA	TOTO	CTCA	) AC 1	1,44C.44C	מערכר איני	C CC		accide	GIT	TCAC	CGT	GTTA	NGCCAGG NGGGATT	
10	ACA	GGCG	TGA	GCC	TOP	יאר (		א א טישי	.G CC	2000	TUAU	י בכיו	CCCA	AAG	TGCT	TGGGATT TTCCCC	9556
	TGT	TA A TIC	ረጥጥል	CTAC	LACCT	י חיחיי		LONA	C 10	) W W L		TAT	ACTG	GTT	CATC	TTCCCC	9616
	ስጥባ ሊጥባ	ישרא כ	ነ ተ ተ ተ ተ	ACTT		. T.T.	CAMO	11111	6 66	TATE	GATT	ATT	TCTC	ATT	TATA	CATTAG	9676
	CTA	CACA	CCT	CCAC	מטעע.	CC C	CMCCC	CCCA	C AG	CTTA	IGGG1	CTC	TTCC	TAA	ATTG	TTATAT	9736
	CVC		NCT.	TCMT	10 v w v	33 (	71600	AATA	6 66	GAAC	TAGT	TTA	TACT	TTC	ATCA	ACTTAG	9796
	שמע	CACA	ACI	TGII	CATA	AA (	SAACA	AAGG	T CA	AGAG	TTAT	GAC	TACT	GAT	TCCA	CAACTG	9856
15	227	CAGA	MOI	1004	COM	AC (	CCGI	GACC	T CT	GCCA	TCCA	GAG	TCTT	TCA	GGCA	TCTTTG	9916
	CAN	TOTAL	MAG	MAAI	GCTA	TT T	TAAT	TTTG	G AG	GTTT	CTCT	ATC	AGTG	CTT	AGGA	TCATGG	9976
	CAT		CAA	COMM	ATGA	.GG (	CAAA	ATTA	A GT	CCAA	AACA	TCT	ACTG	GTT	CCAG	GATTAA	10036
	CAI	COAA	GAA	CCTT	AGGT	GG 1	GCCC	ACAT	G TT	CTGA	TCCA	TCC	TGCA	AAA	TAGA	CATGCT	10096
	GLA	CTAA	CAG	GAAA	AGTG	CA C	GCAG	CACT	A CC	AGTT	GGAT	AAC	CTGC	AAG	ATTA	TAGTTT	10156
	CAA	GTAA	TCT	AACC	ATTT	CTC	CACAA	GGCC	C TA	TTCT	GTGA	CTG	AAAC	ATA	CAAG	AATCTG	10216
20	CAT	TIGG	CCT	TCTA	AGGC	AG C	GCCC	AGCC	A AG	ĢAGA	CCAT	ATT	CAGG	ACA	GAAA	TTCAAG	10276
	ACT	ACTA	TGG	AACT	GGAG	TG C	TTGG	CAGG	G AA	GACA	GAGT	CAA	GGAC	TGC	CAAC	TGAGCC	10336
	AAT	ACAG	CAG	GCTT	ACAC	AG G	BAACC	CAGG	G CC	TAGC	CCTA	CAA	CAAT	ጥልጥ	TGGG	ጥርጥልጥጥ	10396
	CAC	TGTA	AGT	TTTA	ATTT	CA G	GCTC	CACT	G AA	AGAG	TAAG	CTA	AGAT	TCC	TGGC	ACTTTC	10456
	TGT	CTCT	CTC	ACAG	TTGG	CT C	AGAA	ATGA:	G AA	CTGG	TCAG	GCC	AGGC	ATG	CTCC	CTTACA	10516
25	CCT	GGAA	TCC	CAGC	ACTT	TG G	GAGG	CCGA.	A GT	GGGA	GGGT	CAC	ጥጥርል	CCC	CACC	እ ርጥጥር እ	10576
23	GGA	CCAG	CTT	AGGC	AACA	AA G	TGAG	ATAC	c cc	CTGA	CCCC	TTC	TCTA	CAA	AAAT	ጥጥጥፈፈፈ	10636
	TWW	MAAT	TAG	CCAA	ATGT	GG 1	じじいし	TATA	C TT	ACAG	TCCC	AGC	ጥልርጥ	CAG	GAGG	CTCACC	10696
	CAG	GGGG.	ATT	GCTT	GAGC	CC A	GGAA	TTCA.	A GG	CTGC.	AGTG	AGC	בשתמים	ል ጥጥ	ጥር ልር	CACTCC	10756
	ACT	TUTG	PCJ.	GGGC	AACA	GA G	CGAG	ACCC'	т ст	CTCA.	AAGC	ΔΔΔ	4244	444	4404	A CTA CA	10816
	ACT	MGCC.	LAA	GTTT	GTGG	GA Ç	GAGG	TCAT	C AT	CGTC	TTTA	GCC	GTGA.	A TC	CTTA	ם מידי מידי די	10876
30	ACC	ACAG	MAA	TTGA	CATT	AG C	CCAA.	AAAG	C TT	GTGG'	TCTT	TGC	TGGA.	ТОД	CTAC	ግጥል ልጥጥ	10936
	TIG	AGCA.	AAT'	GTGG.	ACAC	CA C	TCAA	TGGG	A GA	GGAG.	AGAA	GTA.	ልርርጥ	CUT	ጥር አጥ	CTATAC	10996
	CCC	AAAA	CTA	GAGG	CCTG	GA A	CTGA.	ATAT(	G CA	TCCC.	ATGA	CAG	GGAG	ጥልል	AGGA	ርልጥጥርር	11056
	GAG	LIAA	JAA I	GGAG.	AGGA	GG T	CAGT.	ACTG	C TG	TTCA	GAGA	ጥጥጥ	րահարա	ייעי	CTAA	CTCTTC	11116
	AGA.	AGCA	AAA	CTAC'	TTTT(	GT T	CTGT'	TTGGʻ	r aa'	TATA	CTTC	AAA	ACAA	ነርጥ ነ	ጥር ልጥ.	ጋጥ ውጥ ውጥ	11176
	AAA	TTGT	rca '	TGTC	CTGA	AA T	AATT	AGGT	A AT	GTTT	TTTT	CTC	TATA	G GA	A AT	G AAT	11233
35																t Asn	
														85			
	CCT	CCT	GAT	AAC	ATC	AAG	GAT	ACA	AAA	AGT	GAC	ATC	ATA	TTC	ጥጥጥ	CAG	11281
	Pro	Pro	Asp	Asn	Ile	Lys	Asp	Thr	Lvs	Ser	Asp	Tle	Tle	Pho	Pha	Glu	11201
			90			•		95	-1-				100	1110	1 110	O,Tu	
	AGA	AGT	GTC	CCA	GGA	CAT	GAT	AAT	AAG	ልጥር	CAA	ጥጥጥ	200	T CT	mc »	mc »	11220
40	Arg	Ser	Val	Pro	Glv	His	Asn	Aen	Tye	Mot	Cla	Dho	Clu	Com	Com	Con	11329
	_	105			0_1		110	ASII	цуз	MEC	GIII	115	GIU	ser	ser	ser	
	TAC		GGA	TAC	ጥጥጥ	СТА	CCT	mCm	CAA	222	CNC	112	010		mm		
	TVI	Glu	Glv	Tyr	Phe	Leu	Ala	Cvc	Clu	T	Class	AGA	GAC	CTT	TTT	AAA	11377
	120		1	- 1 -		125	VI.	Cys	Giu	rys	120	Arg	ASP	reu	Pue		
		ATT	<b>TTG</b>	AAA	444	GAG	CAT	CAA	mm∼	ccc	130	NC N	mam			135	
45	Leu	Ile	Leu	Lys	Lve	Gli	DAT	GI	TOU	C1	GAT.	AGA	TCT	ATA	ATG	TTC	11425
				<b>-</b> 73	140	<b>914</b>	vəb	GIU	ned	OTA	ASD	игд	ser	rre		Phe	
	ACT	GTT	CAA	AAC	GAA	GAC	ጥል ርረ	יייה אַ יייי	י ממי	145	n	~ ~			150		
	Thr	Val	Gln	Asn	Glu	Aen	1760	- I W I ]	nn t	WII.I	CAT						11464
				155.		vaħ											
					•												

13. The genomic DNA of claim 1, which has a nucleotide sequence selected from the group consisting of SEQ ID NO: 14 and its complementary sequence;

55

## SEQ ID NO: 14:

	እርጥጥርርርጥ <del>በ</del> እ	A A A C C TITUTO C C	8 M 8 C C M 8 C 8 C	3 3 C 3 TT T 3 C 3 TT	መል አመመውረርመው	GCTCACATCT	60
5							120
						TAAGACTGAT	
						GCTTCATCTT	180
						GAGGACAGCT	240
	GTGGACTATC	TGGCACTGGA	GACTAAATAA	AGAAAGCAGG	TACAGTCAAT	AAGATCTTCA	300
	GGACATATAC	ATTTTGTTTA	TTAAGAAAAA	GCAAATAAAA	CATTTTTCAG	AAAAAGGCAA	360
	ACATGCTAGA	AAGCATATGA	CTTAGTCATT	TGAGTTTTTA	TTATTAAGGA	AATTTACAGG	420
10	CCCAAGAAAC	ACCTTGCTCA	ATATATTAAA	TTTTATTTTG	GTTTTCAACT	AGACTTTGCT	480
70	TTTCATTTGT	TTGTTTTTGT	GACAAGTTCT	CGCTCTGTCA	CCTAGGCCAA	AGTGTAGTGA	540
						CAGACTCCTG	600
			CATTCCACCA				660
			GCAGCGTTGC				720
			CCCAAAGCAC				780
15							
13						CTGGTATCTT	840
			AGCAATGAAG				900
						CCAAATGCAA	960
						AAAGGTCAGT	1020
						ATCCTATGGA	1080
20	TCATAAAGGC	ACAGACTCAC	TTCTGTCTCT	AGATCTCAAG	CTACCAAAAA	GAAATCTCCC	1140
20	AATAGTTTCT	TGGAGGCCTA	TACTTAGTGA	AAAAGCAGCT	GGAATCAACA	TAGTTCCTCC	1200
	TATGTTGTAG	GACAATCCTA	GCTCTGGGCA	TACGAATACA	TTAAATCCCA	СТТАТСТАТА	1260
	GAGCTTTCTT	AAAGGGAAGA	AATTTGAGTA	GTATGTAAAA	CAGAATAAAA	GATTAAGGCT	1320
	CCATAGGCAT	ACAGCTTACC	TCCAATTCTC	TTGGCCTCTT	GCAATTTCTA	TTATCAGGCT	1380
			ATATTCCGAA				1440
25		-	TGCTACAAGC				1500
			ACTGTCATTA				1560
			ACTGCTATTT		-		1620
			TCATTGTTCC				1680
			GTGCAAGCCT				1740
			GTCTCTCTGG				1800
30			GCCGGGCGCG				1860
							1920
			CACCTGAGGT				
	-		AAAATCCAAA				1980
			CTGAGGTGAG				2040
						GACTCCACCT	2100
35 40			CTATCTTCAC				2160
	TTCATCTAAT	TCAATCTCAT	TTTTTACAAG	TGAGAAAACA	GGGACAGTGA	CGGTGGATCA	2220
	AGTGACACCA	GTAAGACTGA	GCTAAATTAG	AACCGAGATC	TCACTCGAGT	CTGAGGTTAT	2280
	TCCCACTGTC	CAACCTTACT	TTAAAGTAGC	TTCAAATTTT	ACTTTTACTT	TTCCATAAAT	2340
	TCGGAAGGGA	TTTTCCCTAG	GAGTCCAAAT	GTTGAAACCT	GGAAGGGTAT	AGTCTCTGTG	2400
			CTGTCCATAT				2460
			TAACTTTAAC				2520
			AGCAACAACA				2580
			TGCCAGTACC				2640
			AATACTACAC				2700
			AGCCAACAGG				2760
			CCCCTTCTCT				2820
45		·		101010000		TOCHCCIAIG	2020

50

TCAAACGAAA ACTTATAATC AAGTAACTGT TTCTGCAAAA ATAAGTTCGT TTTCCTGTCA TGGCTCAAGG CCTCAGCAGA TCCAGGCCTG GTGGACGGGC TGGTCTTCGT CGTGTGCCAA 5 ACACTGACCA CTGCCCTGGC TCTGCCATCT TAGGCTTAGT GACCTGGCTG TTACTAAGCA CTUTCCCCTC TGCCCCATGC AGCTGTCTCC TTCTAGTCTT CTCCCTCTTC TCAACGCGAT 3060 CCTAGECCCT CAGGCCATTT CACCTCCATT TTCCCTCACT TCCCGCCGCC CCTCCGCACT 3120 TCCTCCTAC TGTTGTTTCC GCCCCACTAG AGCCCCTCAG AGAAAGTTTC CATCCTCGCA CCCTTCCTTG TGTCACAGCC CGTCACATTC TCACAGGCGC CCATCCCTCC AGCCCCACCC CAAGGCCAAT GTACTTCGCG GTATGGGGAC CTTCCTCGTC AGCGAACGCG AGGGAGTGAA 3300 GACCCTGGGC GCGGGTGCT CGGACTTCGG GGGTGGAGGT GGGAAGCGCG CCGCACTCCC 3360 AGCAGCCCCT GCACGAGTCA CGTGACAGCT CTCCCACCAC CACCCCCCC AACTTCCCCA 3420 CCGTAGCCTC CCAGAGCCAG GCCCCACGGA AAGGCAGCTT TTTCCCGGTT TTCTCCCGCT 3480 CTTTCCCCTC CACTTGGAAT ACTCGTGAAA CAAAAATCTC TCCCTGCCAC CCTGTGTGTG TTTGAACCAG GAAAAAATCT GAAACTGGTC AAGAAAGAAC AAGGAAGACT TGCCAAAGCA 15 AGGCCGGTGT GTGTCCCAGC AGCTTAGAAT CTCAGCAAAG GAACACAAAA TAGCACATCC 3660 ACGGCCTCTT TTCGAGTAAA ATTTACTTGG TTTGTTTGCA GGAAGGGTTT AAAACTGCGT 3720 TTGCAGATGC TCTGTTTGCA GGAAGGCTTT AATCACGTGT TCCCCTGGCC CACAAGCAAG GCTTTTAGAT CCAGAGCCTC AGTTACTGCC CCCTCTTCCT CTTTGGTGCA ACCAAACGTT 3780 CAGAATCACG CCTTCTTAGA AAATTCTTAC CCCGGGTGTG TCAATAAGTT AAGTCTAATT 3900 GGCAACAGCT ATCAAAAAGT GTTGCATAAC ACACATGCCT CACATAATTG TAGCTTTGCC 3960 20 TCATCGGGTG TTTTAATGCG GAGGCTTTGA CCTGCAATTT CAAAGATATA CATTCCAAGC 4020 TTACGCCCAG TTAGTGGATG TGGAAGAAAA AAAAAAGCAA ATTACCTCAT AACACAAAGG TCAATAACAC ACATCCATAA GCTCCAGGTA CAAAATCTTA CATCTTAGAG AACTATATTT AACATTTACA TACATTACTA AGGTTTTTTT TTTCCTTTTG CTTGATTAAA TGTTAGTTAT 4200 CATTAAGTCT TGGAATTATT CTGTGTGTCT ATATTTATTT GCTGTTTGTG AAGAAGCCGG 4260 25 TTGTTTTAAA TAAGTTCCTA GAAAATAAGC GCTCAATGTG TTTAATCTGA GTTGCTAATA TTGTGAAATA TAGGCCACAT AATACTAGCC TAGATAACTA TGGCGAAGTA AGGAGTCTCA 4320 4380 AACACTGTCC CAGAACAATA GCAATCTGTG TTGAATTTTT ACCCTCTGTG GTAAAATGAA 4440 GGGAAAAGGA ATGAAGTTTT AGTTTGCCTT AATTTTTATC TTTATTGTTT CAGACTCTTC 4500 AGCAGTATAA AGTTTTCATC AAGTCAAATA TATTCACTTT AAAGTGACTG TGCTTTATTC 4560 TGATACCATC TCCTTCCTAA TTTGGGGGGC CAGCTGAGAT AAGTTTTATG AAATAAAAAG 4620 30 ATTAAAAATT CTTACATTTT TAGTGTCCTT CCTTGGTAAA ATGTAGAGTT GTCCACTGTG 4680 TTTATCTCCT CCTCCTTATT ATCATGGTTG CTGTTATTAT TTTTAATGGT TCATTAAACC CAAGGGTCTG GGAAATACTC ATGGAATTCA TCTCACAGCC TTCACACTGT ATGATATTA 4740 4800 AACAGGTGGT TGTCCATCTG ATTCTTAAAA TATTTCCAAG AAAAATGATT CCACCTAATG 4860 4920 AAATATTAAC TTCCATTGCA TAAGCTAAAT GGGTAGGAAT AAGTGAGATG ATATTGTTAT 4980 35 CTAGAGCTTT AAAATATTCA AAGGGCTGTC ATCATTATCT CATTTAATCT TTGAAAACAA 5040 CTCTATGAAG TACAAAGGAC ACTGAGACAT TTGTTGCTCT ATATCAAAGA AAAAAGTGTT 5100 TGTCCCAAAA CTTCAAAATG TGTAAATTAC ACATTCTGCA TCTTTACAGC TGGAGAAAAT 5160 TCACTGGCAA TGGAATATTT AAAATTAGAG CTTGCTTAGT GTGCTGCTTC TGATCACTAC 5220 TTGATCCCAC TTCGTGCTTT CATGTTAATT GGCCCAATTG GACTCTACAG TTGGAAGGTG 5280 AAAACTTACT ATTTCAACTT GAGTCACGTA TGTATTCTTA TCATATACTT CTTAAAGGTA 5340 CTATTTTTT TCTTCTGATA GTCACCACAC CAAGCACTTC CAGCCACCCT GCCACAGACT 5400 TCCTTTGTAA TCACTGTTGA AGGACATGAT GTTTTTATGA CTTCCCGAAA TGAAAACCCT 5460 ATCTTGTTTT TAAAACAAAC AAACCAACAA AAAGTAGTGT TTATGTAAGC ATTTTGTTCC 5520 CTGACTCTAG GAACCCCTCT GTTTTTATAT CAACTCTGTA CTGGCAAAAC ACAAAAACAA 5580 AATGCCACCT TGCTAATTCC CTTCCTAGCA AAGTAATACA GTTTAGCACA TGTTCAAGAA 5640 AAAAATGGCT AAGAAATTTT GTTTCCACTA ATTATTTTCA AGACTGTGAT ATTTACACTC 5700 TGCTCTTCAA ACGTTACATT TTATAAGACT ATTTTTTAAC ATGTTGAACA TAAGCCCTAA ATATATGTAT CCTTAAATTG TATTTCAAAT ATTTTAGGTC AGTCTTTGCT ATCATTCAG 5820 GAATAGAAAG TTTTAACACT GGAAACTGCA AGTAAATATT TGCCCTCTTA CCTGAATTTT 5880 GGTAGCCCTC TCCCCAAGCT TACTTTCTGT TGCAGAAAGT GTAAAAATTA TTACATAAAA 5940 TTCTAATGAT GGTATCCGTG TGGCTTGCAT CTGATACAGC AGATAAAGAA GTTTTATGAA 50 AATGGACTCC TGTTCCACTG AAAAGTAAAT CTTAATGGCC TGTATCAACT ATCCTTTGAC 6060 ACCATATTGA GCTTGGGAGG AAGGGGAAGT CCTGAATGAG GTTATAAAGT AAAAGAAAAT ATTTGCAAAA TGTTCCTTTT TTTAAAATGT TACATTTTAG AAATATTTTA AGTGTTGTAA 6180 CATTGTAGGA ATTACCCCAA TAGGACTGAT TATTCCGCAT TGTAAAATAA GAAAAAGTTT 6240

TGTGCTGAAG TGTGACCAGG AAGTCTGAAA ATGAAGAGAG ACAGATGACA AAAGAAGATG CTTCTAATGG ACTAAGGAGG TGCTTTCTTA AAGTCAGAAA GAGATACTCA GAAAGAGGTA 6360 CAGGTTTTGG AAGGCACAGA GCCCCAACTT TTACGGAAGA AAAGATTTCA TGAAAATAGT 6420 GATATTACAT TAAAAGAAGT ACTCGTATCC TCTGCCACTT TATTTCGACT TCCATTGCCC 5 TAGGAAAGAG CCTGTTTGAA GGCGGGCCCA AGGAGTGCCG ACAGCAGTCT CCTCCCTCCA CCTTCTTCCT CATTCTCTC CCAGCTTGCT GAGCCCTTTG CTCCCCTGGC GACTGCCTGG ACAGTCAGCA AGGAATTGTC TCCCAGTGCA TTTTGCCCTC CTGGCTGCCA ACTCTGGCTG CTAAAGCGGC TGCCACCTGC TGCAGTCTAC ACAGCTTCGG GAAGAGGAAA GGAACCTCAG ACCTTCCAGA TCGCTTCCTC TCGCAACAAA CTATTTGTCG CAGGTAAGAA ATATCATTCC TCTTTATTTG GAAAGTCAGC CATGGCAATT AGAGGTAAAT AAGCTAGAAA GCAATTGAGA 10 GGAATATAAA CCATCTAGCA TCACTACGAT GAGCAGTCAG TATCAACATA AGAAATATAA GCAAAGTCAG AGTAGAATTT TTTTCTTTTA TCAGATATGG GAGAGTATCA CTTTAGAGGA 6960 GAGGTTCTCA AACTTTTTGC TCTCATGTTC CCTTTACACT AAGCACATCA CATGTTAGCA 7020 TAAGTAACAT TTTTAATTAA AAATAACTAT GTACTTTTTT AACAACAAAA AAAAGCATAA 7080 AGAGTGACAC TTTTTTATTT TTACAAGTGT TTTAACTGGT TTAATAGAAG CCATATAGAT 7140 CTGCTGGATT CTCATCTGCT TTGCATTCAG ACTACTGCAA TATTGCACAG AATGCAGCCT 15 CTGGTAAACT CTGTTGTACA CTCATGAGAG AATGGGTGAA AAAGACAAAT TACGTCTTAG 7260 AATTATTAGA AATAGCTTTC ACTTTAGGAA CTCCCTGAGA ATTGCTGCTT TAGAGTGGTA 7320 AGATAAATAA GCTTCTCTTT AAACGGAATC TCAAGACAGA ATCAGTTACA TTAAAAGCAA 7380 ACAAAAATT TGCCCATGGT TAGTCATCTT GTGAAATCTG CCACACCTTT GGACTGGGCT ACAATTGGAT AATATAGCAT TCCCCGAGAT AATTTTCTCT CACAATTAAG GAAAGGGCTG 20 7500 AATAAATATC TCTGTTTGAA GTTGAATAAC AAAAATTAGG ACCCCCTAAA TTTTAGGGCT 7560 CCTGAAATTC GTCTTTTTGC CTATATTCAG CTACTTTACG TTCTATTAAA TCTTCTTTCA 7620 GGCCAGGTGC ACTAGCTCAT GCCTAGAATC TCAGGCAGGC CTGAGCCCAG GAATTTGAGA CCAGCCAGGG CAACACAGTC TCTACAAAAA AATAAAAAAT TACCTGGGTG TGTTGGTGCA 7680 TGCCTGTAGA ACTACTCAGG ATGCTGAGGA CTGCTTGAGC CCAGGATAGC CAAATCTGTG GTGAGTTCAG CCACTAAACA GAGCGAGACT TTCTCAAAAA AACAAACAAACA 25 7860 AACTTCCTTC AAAATAACTT TTTATCTGCA ATGTTTTCCT ATTGCCTGTG AGATTAAATT TACTCTTTTA CCTGATTTCC AAAGCCCTCC ATAATCTAAT CCGACTTTAC CTTGTGTTCA CTGCAAAATA GCAGGACTGT TCCACTACAA TCCAAAAATC ACAGGTTGGG TGCAGTGGCT 8040 CACTCCTGTA ATCCCAACAC TTTGGAAGGC CAAGGCAGGT GGATTGCTTC AGCTCAGGAG 8100 TTCAAGACCA GCCTGGGCAA CATGGCAAAA ACCCTGTCTC TCCAAAACAT ACAAAAATTA 8160 GCCAGATGTG GTAGTATGTG CCTGTAGTCC CAACTACTCA AAAGGCTAAG GCAAGAGGAT 30 8220 CACTTGAGCC CAGGAGGTCA AGGCTACAGT GAGCCATGTT TACTGTGTCA CTGCACTCCA 8340 ACATCGCTCT ATTCAGTTCA CCCCCACCAC AACATTGTTT TGATTATCAC ATAAATGCTG 8400 GTCCATTGCC TTCTCTATCT ATTCAAATCT TTAAGCATTC TTTGAGATTC AACTCAATTC 8460 TCCTTTTCAA ACTAGGCCAT TTAAACTACA TCAGTTCCAT TTTGATTTTC TTGCTTTGAG 35 8520 TCTACAGACT CAAAAACAAA AACTTAAAAA CTTATTTTTT AAGTTTTCTG CTACTCTCAC 8580 TTCTTCAACA CTCACATACA CGCATTCATA ATAAGATGGC AGAATGTTCA AGGATAAAAT 8640 GATTTATAGA ACTGAAAAGT TAGGTTTTGA TCTTGTTGCT GTCAAGATGA CTACCTACCT 8700 GATCTCAGGT AATTAATTAT GTAGCATGCT CCCTCATTTC ATCCCATACC TATTCAACAG 8760 GATTGGAATT CCACAGCAAG GATAAACATA ATCATAGTTG CTTTTCAAGT TCAAGGCATT TTAACTTTTA ATCTAGTAGT ATGTTTGTTG TTGTTGTTGT TGTTTGAGAT GGAGCCCTGC 40 TGTGTCACCC AGGCTGGAGT GCAGTGGCAC GAACTCGGCT CACTGCAACC TCTGCCTCAT GGGTTCAATC AGTTATTCTG CCTCAGTGTC CCAAGTAGCT GGGACTACAA GGCACATGCC ACCATGCCTG GCTAATTTTT GTATTTTTAG TAGAAACAGG GCTTCACCAT GTTGGCCAGG 8940 9000 CTGGTCTCGA ACTCCTGACC TCAAGTGATC CAGCCGCCTC GGCCTCCCAA AGTGCTGGGA 9120 TTACAGGCAT AAGCCACCGT GCCCAGCCTA ATAGTATGTT TTTAAACTCT TAGTGGCTTA 9180 ACAATGCTGG TTGTATAATA AATATGCCAT AAATATTTAC TGTCTTAGAA TTATGAAGAA 9240 GTGGTTACTA GGCCGTTTGC CACATATCAA TGGTTCTCTC CTTACAGCTT TAATTAGAGT 9300 CTAGAATTGC AGGTTGGTAG AGCTGGAACA GACCTTAAAG ATTGACTAGC CAACTTCCTT 9360 GTCCAAATGA GGGAACTGAG ACCCTTAAAA TTAAGTGACT TGCCCCAGAC AAAACTGGAA 9420 CTCATGTGTC CTAATTTCCA TCATGAAATT CTACCATTCA CTAGCCTCTG GCTAGTTGTC ANAGTATTGC ATAACTAAAT TTTTATGTCT GTTTTAAAGA ACAAATTGTC ACTGCTTACT 50 9540 CCTGGGAGGG TCTTTCTGAG GTGGTTTATA ACTCTTAAAA AAAAAAAAGT CAGTAGTCTG 9600 AGAATTTTAG ACGAAATAGT CAAAGCATTT TTATCCAATG GATCTATAAT TTTCATAGAT

	TAGAGTTAAA	TCAAAGAAAC	ACGGATGAGA	AAGGAAGAGG	AAAATTGAGG	AGAGGAGGAA	9720
		AACACACTAC					9780
5							
J		ANTAAGAATG					9840
	AGGGAAAAAT	GGAGAAGACA	TTAGAAAAAT	TATTCTATTT	TTAAAATTCT	GTTTTCAGGC	9900
	TTCCCTCCTG	TTCTTCCTCC	TTCTCATTGG	TTTTCAGGTG	GAGGGAAAGT	TTAAGATGGA	9960
	TATATATAT	ATATTCTACA	CATCCCTTTC	TACGCTGTTG	TCATGCCAAC	AAGGTTTATC	10020
		TTTATTCATA					10080
10	AGGTGTTGAA	AATTCAGGGG	AAAAAAGACA	ACTCATTGTC	TTAAAACTCA	GATGAAAGCT	10140
10	GAACAGACCT	ATTTTTAATC	AAAGTAATCT	CAATTTAGGG	TAGTAAGAGC	TATTTAAGAA	10200
	GCATGAACAG	GTGTGAAGGA	GGTAGGACTC	TGAGGAGAGA	ATAGTTAGCT	AGGAATGAAA	10260
		GTTTTCCTAG					10320
		CAGGCAAAAA					10380
	AGAGAATAAC	TGTGGCAACA	ATGGAGGAGA	GTCTGGAAGC	AAGAAAACCA	AGTAGAAGAG	10440
	TATTAAAATA	GAAGATGCCA	GGGGTAATGA	GGGCTTGATT	TAAAACAGTG	CTGTTGGAGA	10500
15	TCGAGAGGAG	ATACCAAATT	CTGGAGACAT	TTCTGAGTTA	GAACCTACAG	TATTTATCAG	10560
		GATTAGACAA					10620
		ATGTTTTGAA					10680
		TTAAAATTTG					10740
	CCCTAACTCT	CTACTTTGTA	GCCAGACTTC	CTAAAAGAAT	AGTTTGTAGT	CACTGTCTTT	10800
	ACTTTTCCCC	TCCCATTCTG	TCCTAGATAT	TTGTCCACCT	ACCATCTGCT	GCCTCCACTT	10860
20		GTTCTACGGT					10920
		TGTAAATTAG					10980
		CGACTTCTGT					11040
	ACTTGAGCTG	TATTTTCGTT	CTTCTTTCTT	GATGAATGAG	GTAACCACTC	AACAAATTGC	11100
	CCAAGCCAAA	AACTACGAAG	TCATCCTCAG	TTCCTCCTTC	TTCTGTTTGA	CCCACAACAG	11160
0.5		AAATCCCGCT					11220
25							11280
		TAATTGTTAA					
		GTCCATTCTT					11340
	TTTGCTGCCC	AGGCTGAAGT	GCAGTGGAGC	ACTTCACTGC	AACCACAGCC	TCCTGGGTTT	11400
	AAGCAGTTCA	CCCTCCCGAG	TAGCTGGGAC	TACAGGTATG	TGCCACCACA	CCCAGCTAAT	11460
		CAGTAGAGAC					11520
20		ATCCACCCAC					11580
30							
		CCAGATCCAT					11640
	TTGACCATAT	CTTTCTCCAA	TTTAAGTCAG	TATTTTTTT	TTCAGGAAAA	AACAGTTCAA	11700
	ACTCTTTAGT	CTGCTTACAC	AAGGCCTTTG	TAGTCTGACT	CTTCTTTCCA	AGCTTTCATC	11760
	AAAGTATACT	GCAAGTTACA	TTTTATGTGA	ATTGAATTAG	GCAACGGTAT	AAAAATTATA	11820
		GCAAAATGGA					11880
35							11940
35		GCTGTCAGGT					
		TACTTGGTAG					12000
	TAAATAACTC	AGAAGTTCTT	CAGACATACA	GGTTATTATT	GTGCTTTTTA	AACATAATTT	12060
	TAAATAATTT	TATATATGAT	AATGTTATCC	AAGTGCTAAG	GGATGTATTG	TTACTGCTGT	12120
	CCAAAAAAAA	AAAAAAAAA	AACTCCAAAT	AAATATGTTG	AAACCAAGTT	TATATGCAAG	12180
		TAAAAAGGCC					12240
40							
40		GTAATAATGC					12300
	GGTTGTTGTA	GGTTTATGTA	CTCCAAGAAT	ATCTCCTCTC	AAACTTTTAC	GTTTTTTCCA	12360
	AAGAAAAGTT	AACTTTGGCT	GGGCGCAGTG	GCTCTTGCCT	GTAGTCCCAG	CCTTTGGGAG	12420
	GCCAAGGCGG	GCAGATCACC	TGAGGTCAGG	AGTTTGAGAC	CAGCCTGACC	AAAAATGGAG	12480
		CCCTCACTAC					12540
45	-	AGCACTTTGG					12600
40		GGAGAAACCC					12660
	CATGACTGTA	ATCCCAGCTA	CTCAGGAGGC	TAAGGCAGAG	AATCACTTGA	ACCCAGGCAG	12720
	TGGAGGTTGC	AGTGAGCCGA	GATCGTGCCA	TTGCACTCCA	GCCTGGGCAA	CAAGAGCGAA	12780
		CAAAAAACAA					12840
		CTTTCTGTAA			-		12900
50		AACCCATATT					12960
J-		GGAAGCTCTT					13020
	ACAGTCATCA	CTTTCTGTGG	TAACCTTCTC	CAGCACCATC	AAACAGAAAG	AATGAATCTC	13080

```
TTGTAAATTC AGCTCTTACG TCATTCATTA CATTATTTTG TAACTCTTTA TAGATTCTTC 13140
        TCTCCCACTA GACTCTGAGT CACTGGAGAG TAGGAGCCAA CTCTCATTCA TGTGTGGTTT 13200
        GGTCAGCTAC TGGCCACATT CCTGATGCAT AGTTAATGCT CAAACCTTAA CTGGTGAATC 13260
5
        AGCTCAAATA TTGTCCTTCT CTAAATCCAT TCACTCATTG ACTAACTATG TACTCAAAAT 13320
        AGTAAACACC AGTAATTTAA TCCAATTCCT GCCCATACTG CTTGGTACAT TTCAGGTGAA 13380
        TTAGTTTGAT AAATATGTGT GTATTACATA ATATTAAAGT ATGTACAGAA GATCATGCTA 13440
        ATCATAATTC ACAACTGATA ACTAATCAAA CATAAATGCT CTCAGGTTAA CAAATGTCTG
        CCTTCTCAGT TAATGCAGTC ATTAACAAAC ACCTTCTGAT GCTGATAATA GGGCCTTGTT
                                                                                   13560
        CAGCAATGAA GCCATAAAGG TGAATAAAGA ACATGCCCTC GTGGAGCTCA CAGCCTAGTC 13620
        ATTATTGTTC TGATTTTAA TATTAATGTT GGTTTGGGTT TTGGTGAAAA ATGTTTAGAC 13680
        TTATCTTAGT GATCTTTCA TCCTTTGCTA TATTATTTT CTCTAAGAGT CTTCCTTATC
        CCCTCCTTTA AAAAACTAGG TGATAATTCT AAATTGTAAA TTTAAATATT ATAAATAGCT
        TATAAAATT AATATTATA ATATTTAAAT GTTTGATAAA TATTTAAATT TTATAATATT
        TARATGTTTA TTTARATTCA TTTGTACATC AGTTTTTATT TTATTTARAT GTGTTGGCCA
15
       GGCATGGTGG CTGACACCTA TAATCCCAGA ACTTTGAGAG GCCAAGTCAG GCAAACCATT
                                                                                   13980
        TGAGCTCAGG AGTTTGAGAC CACCCTGGGC AACGTGGTGA AACCCTGTCT CTACCAAACA
                                                                                   14040
        TATGAAAACT TATCTGGGTG TGGTGGCACG CATCTGTGGT CCCAGATGGG AGTCCCAGGC
                                                                                  14100
        TAAGATGGGA GAATCGCTTG AACCCAGGTG AGAGGGGTGG GGTGGATGTT GCAGTGAGCT
       14220
                                                                                   14280
20
       TAAACCTAGA TAAAACTATC AAATAAGGCC TGGGTACAGT GACTCATGCC TGTAATCTCA
                                                                                   14340
        AGCACTTTGG GAGGCCAAAA TTATACAAAG TTAGTTGTAT AACACCAACT AACAACTATT 14400
       TTGGGGTTAG CTTAATTCAG ATTAATTTTT TTTAAACTGA GTTTTAAATT CCTGCTTACT 14460
        CTACCATACA TGCTAGGCCT CATATTATGC TAGAAAAATT TTGAGCACAG ATTTATGAAT
       ACTUTCUTGU ATACCATTTA ATTTTTAAAC AAATTTTAAT GCAGTATATA TGTGCCTTTT
TACCAACACA TTAAATAATA AGATCTACTG TGAGGACTAA ATTTCTGTAA TTTCAAAGTA
                                                                                   14580
25
       GTAATGAGTT TAAACCATGT CTCAAGATCT CTGCAATAAC TGTAGCACAA CAGAAAATAG
                                                                                   14700
        STATTTCTAT TAATGACAGA GTCACAAGTA CTACTAATAA TACTGTGGTT TGTTTCCTGC
                                                                                  14760
        AACTAATCAT GGGAGGAATG CTAAATTTCA GAGGTTGGTG AAAATACATG TGTATTTTTT
        TCCCCATCCA AGTTCACAGA TTTCTCACAC TGAGAACTCC TATTCCATAA CAAAATTCTG 14880
       GAAGCCTGCA CACCGTATTG GAAGAAGGC AGAAAGGAAA AGCAAATGGA AGGATTTAAA 14940
TTTTTTTCAA ATCCTGTATC CCTTGATTTT ACAGCAAGAT TGTATTTATG TATTACTTGT 15000
GTTAAAAAATA TAGTATAATC GAGACTCCAG ATCAAAAATC ACCGCAGCTC AGGGAGAAAG 15060
30
                                                                                  15060
       AGGGCCACCA AATGCCAGAG CCCTTCAGCC TTCTCCCACC CTGCCTGTAC CCTCAGATGG 15120
       AAGCACTTTT TTATCATTGT TTCACCTTTA GCATTTTGAC AATGAAGTCA CAAACCTTCA 15180
        GCCTCTCACC CATAGGAACC CACTGGTTGT AAGAGAAGGA TGAAGCCAGT CCTTCCTAAA
       GGGCACGATT AGATGTGTTT ATGGCATCCT CAGGTGAAAC TATATTTATA TTGACAATAT ATTTATATTT CTCAAGGAAT ACTAGAATAA TGATTCAGTT CAGTACTAGG CCATTTATCT
                                                                                  15300
35
       ACCCTTTATA ATATTGTTTA ATGAGAAAAT GCTTTCTATC TTCCAAATAT CTGATGATTT 15420
       GTANGAGAAC ACTTAAACAT GGGTATTCAT AAGCTGAAAC TTCTGGCATT TATTGAATGT 15480
       CAAGATTGTT CATCAGTATA CTAGGTGATT AACTGACCAC TGAACTTGAA GGTAGTATAA 15540
       AGTAGTAGTA AAAGGTACAA TCATTGTCTC TTAACAGATG GCTCTTTGCT TTCATTAGGA 15600
       ATAAAG ATG GCT GCT GAA CCA GTA GAA GAC AAT TGC ATC AAC TTT GTG GCA 15651
40
               Met Ala Ala Glu Pro Val Glu Asp Asn Cys Ile Asn Phe Val Ala
                   -35
                                          -30
                                                                 -25
       ATG AAA TTT ATT GAC AAT ACG CTT TAC TTT ATA G
                                                            GTAAGGC TAATGCCATA 15702
       Met Lys Phe Ile Asp Asn Thr Leu Tyr Phe Ile Ala
                                  -15
                                                         -10
       GAACAAATAC CAGGTTCAGA TAAATCTATT CAATTAGAAA AGATGTTGTG AGGTGAACTA 15762
       TTAAGTGACT CTTTGTGTCA CCAAATTTCA CTGTAATATT AATGGCTCTT AAAAAAATAG 15822
TGGACCTCTA GAAATTAACC ACAACATGTC CAAGGTCTCA GCACCTTGTC ACACCACGTG 15882
TCCTGGCACT TTAATCAGCA GTAGCTCACT CTCCAGTTGG CAGTAAGTGC ACATCATGAA 15942
       AATCCCAGTT TTCATGGGAA AATCCCAGTT TTCATTGGAT TTCCATGGGA AAAATCCCAG 16002
       TACAAAACTG GGTGCATTCA GGAAATACAA TTTCCCAAAG CAAATTGGCA AATTATGTAA 16062
       GAGATTCTCT AAATTTAGAG TTCCGTGAAT TACACCATTT TATGTAAATA TGTTTGACAA 16122
50
       GTAAAAATTG ATTCTTTTT TTTTTTCTC TTGCCCAGGC TGGAGTGCAG TGGCACAATC 16182
TCTGCTCACT GCAACCTCCA CCTCCTGGGT TCAAGCAATT CTCCTGCCTC AGCCTTCTGA 16242
```

```
GTAGCTGGGA CTACAGGTGC ATCCCGCCAT GCCTGGCTAA TTTTTGGGTA TTTTTACTAG 16302
       AGACAGGGTT TTGGCATGTT GTCCAGGCTG GTCTTGGACT CCTGATCTCA GATGATCCTC
                                                                               16362
       CTGGCTCGGG CTCCCAAAGT GCTGGGATTA CAGGCATGAA CCACCACACA TGGCCTAAAA
                                                                               16422
5
       ATTGATTCTT ATGATTAATC TCCTGTGAAC AATTTGGCTT CATTTGAAAG TTTGCCTTCA 16432
       TTTGAAACCT TCATTTAAAA GCCTGAGCAA CAAAGTGAGA CCCCATCTCT ACAAAAACT
       GCAAAATATC CTGTGGACAC CTCCTACCTT CTGTGGAGGC TGAAGCAGGA GGATCACTTG
       AGCCTAGGAA TTTGAGCCTG CAGTGAGCTA TGATCCCACC CCTACACTCC AGCCTGCATG
       ACAGTAGACO CTGACACACA CACACAAAAA AAAACCTTCA TAAAAAATTA TTAGTTGACT
       TTTCTTAGGT GACTTTCCGT TTAAGCAATA AATTTAAAAG TAAAATCTCT AATTTTAGAA
                                                                               16782
10
       AATTTATTT TAGTTACATA TTGAAATTTT TAAACCCTAG GTTTAAGTTT TATGTCTAAA
                                                                               16842
       TTACCTGAGA ACACACTAAG TCTGATAAGC TTCATTTTAT GGGCCTTTTG GATGATTATA
                                                                               16902
       TAATATTCTG ATGAAAGCCA AGACAGACCC TTAAACCATA AAAATAGGAG TTCGAGAAAG 16962
       AGGAGTAGCA AAAGTAAAAG CTAGAATGAG ATTGAATTCT GAGTCGAAAT ACAAAATTTT
       ACATATTCTC TTTCTCTTT TTTCCCCCTC TTAG CT GAA GAT GAT G
                                                                     GTAAAGT 17075
15
                                                Ala Glu Asp Asp Glu
                                                -10
       AGAAATGAAT TTATTTTCT TTGCAAACTA AGTATCTGCT TGAGACACAT CTATCTCACC 17135
       ATTGTCAGCT GAGGAAAAA AAAAATGGTT CTCATGCTAC CAATCTGCCT TCAAAGAAAT
                                                                               17195
       CTGCACTCAG TAGCACAGCT TTGGAATGAA GATGATCATA AGAGATACAA AGAAGAACCT 17255
       CTAGCAAAAG ATGCTTCTCT ATGCCTTAAA AAATTCTCCA GCTCTTAGAA TCTACAAAAT 17315
20
       AGACTTTGCC TGTTTCATTG GTCCTAAGAT TAGCATGAAG CCATGGATTC TGTTGTAGGG
                                                                               17375
       GGAGCGTTGC ATAGGAAAAA GGGATTGAAG CATTAGAATT GTCCAAAATC AGTAACACCT CCTCTCAGAA ATGCTTTGGG AAGAAGCCTG GAAGGTTCCG GGTTGGTGGT GGGGTGGGGC
                                                                               17435
       AGAAAATTCT GGAAGTAGAG GAGATAGGAA TGGGTGGGGC AAGAAGACCA CATTCAGAGG
                                                                               17555
       CCAAAAGCTG AAAGAAACCA TGGCATTTAT GATGAATTCA GGGTAATTCA GAATGGAAGT 17615
       AGAGTAGGAG TAGGAGACTG GTGAGAGGAG CTAGAGTGAT AAACAGGGTG TAGAGCAAGA 17675
25
       CGTTCTCTCA CCCCAAGATS TGAAATTTGG ACTTTATCTT GGAGATAATA GGGTTAATTA
       AGCACAATAT GTATTAGCTA GGGTAAAGAT TAGTTTGTTG TAACAAAGAC ATCCAAAGAT ACAGTAGCTG AATAAGATAG AGAATTTTTC TCTCAAAGAA AGTCTAAGTA GGCAGCTCAG
       AAGTACTATG GCTGGAAGCA ACCTGATGAT ATTGGGACCC CCAACCTTCT TCAGTCTTGT
                                                                               17915
       ACCCATCATC CCCTAGTTGT TGATCTCACT CACATAGTTG AAAATCATCA TACTTCCTGG
                                                                               17975
       GTTCATATCC CAGTTATCAA GAAAGGGTCA AGAGAAGTCA GGCTCATTCC TTTCAAAGAC
30
       TCTAATTGGA AGTTAAACAC ATCAATCCCC CTCATATTCC ATTGACTAGA ATTTAATCAC
       ATGGCCACAC CAAGTGCAAG GAAATCTGGA AAATATAATC TTTATTCCAG GTAGCCATAT GACTCTTTAA AATTCAGAAA TAATATTTT TTAAAATATC ATTCTGGCTT TGGTATAAAG
                                                                               18215
       AATTGATGGT GTGGGGTGAG GAGGCCAAAA TTAAGGGTTG AGAGCCTATT ATTTTAGTTA
                                                                               18275
       TTACAAGAAA TGATGGTGTC ATGAATTAAG GTAGACATAG GGGAGTGCTG ATGAGGAGCT
                                                                               18335
       GTGAATGGAT TTTAGAAACA CTTGAGAGAA TCAATAGGAC ATGATTTAGG GTTGGATTTG 18395
35
       GAAAGGAGAA GAAAGTAGAA AAGATGATGC CTACATTTTT CACTTAGGCA ATTTGTACCA 19455
       TTCAGTGAAA TAGGGAACAC AGGAGGAAGA GCAGGTTTTG GTGTATACAA AGAGGAGGAT 18515
       GGATGACGCA TTTCGTTTTG GATCTGAGAT GTCTGTGGAA CGTCCTAGTG GAGATGTCCA
                                                                               18575
       CAAACTCTTC TACATGTGGT TCTGAGTTCA GGACACAGAT TTGGGCTGGA GATAGAGATA
                                                                               18635
       TTGTAGGCTT ATACATAGAA ATGGCATTTG AATCTATAGA GATAAAAAGA CACATCAGAG
                                                                               18695
       GAAATGTGTA AAGTGAGAGA GGAAAAGCCA AGTACTGTGC TGGGGGGAAT ACCTACATTT
40
                                                                               18755
       AAAGGATGCA GTAGAAAGAA GCTAATAAAC AACAGAGAGC AGACTAACCA AAAGGGGAGA
                                                                               18815
       AGAAAAACCA AGAGAATTCC ACCGACTCCC AGGAGAGCAT TTCAAGATTG AGGGGATAGG
                                                                               18875
       TGTTGTGTTG AATTTTGCAG CCTTGAGAAT CAAGGGCCAG AACACAGCTT TTAGATTTAG 18935
       CAACAAGGAG TTTGGTGATC TCAGTGAAAG CAGCTTGATG GTGAAATGGA GGCAGAGGCA 18995
GATTGCAATG AGTGAAACAG TGAATGGGAA GTGAAGAAAT GATACAGATA ATTCTTGCTA 19055
45
       AAAGCTTGGC TGTTAAAAGG AGGAGAAAA CAAGACTAGC TGCAAAGTGA GATTGGGTTG
                                                                               19115
       ATGGAGCAGT TTTAAATCTC AAAATAAAGA GCTTTGTGCT TTTTTGATTA TGAAAATAAT
                                                                               19175
       GTGTTAATTG TAACTAATTG AGGCAATGAA AAAAGATAAT AATATGAAAG ATAAAAATAT
                                                                               19235
       AAAAACCACC CAGAAATAAT GATAGCTACC ATTTTGATAC AATATTTCTA CACTCCTTTC
       TATGTATATA TACAGACACA GAAATGCTTA TATTTTTATT AAAAGGGATT GTACTATACC
       TAAGCTGCTT TTTCTAGTTA GTGATATATA TGGACATCTC TCCATGGCAA CGAGTAATTG
                                                                               19415
50
       CAGTTATATT AAGTTCATGA TATTTCACAA TAAGGGCATA TCTTTGCCCT TTTTATTTAA
                                                                               19475
       TCAATTCTTA ATTGGTGAAT GTTTGTTTCC AGTTTGTTGT TGTTATTAAC AATGTTCCCA 19535
```

```
TAAGCATTCC TGTACACCAA TGTTCACACA TTTGTCTGAT TTTTTCTTCA GGATAAAACC 19595
      CAGGAGGTAG AATTGCTGGG TTGATAGAAG AGAAAGGATG ATTGCCAAAT TAAAGCTTCA
                                                                      19655
      GTAGAGGGTA CATGCCGAGC ACAAATGGGA TCAGCCCTAG ATACCAGAAA TGGCACTTTC 19715
      TCATTTCCCC TTGGGACAAA AGGGAGAGAG GCAATAACTG TGCTGCCAGA GTTAAATTTG 19775
5
      TACGTGGACT AGCAGGAAAT CATTTGCTGA AAATGAAAAC AGAGATGATG TTGTAGAGGT 19835
      CCTGAAGAGA GCAAAGAAAA TTTGAAATTG CGGCTATCAG CTATGGAAGA GAGTGCTGAA
      CTGGAAAACA AAAGAAGTAT TGACAATTGG TATGCTTGTA ATGGCACCGA TTTGAACGCT
      TGTGCCATTG TTCACCAGCA GCACTCAGCA GCCAAGTTTG GAGTTTTGTA GCAGAAAGAC
      AAATAAGTTA GGGATTTAAT ATCCTGGCCA AATGGTAGAC AAAATGAACT CTGAGATCCA
      GCTGCACAGG GAAGGAAGGG AAGACGGGAA GAGGTTAGAT AGGAAATACA AGAGTCAGGA
10
                                                                       20135
      GACTGGAAGA TGTTGTGATA TTTAAGAACA CATAGAGTTG GAGTAAAAGT GTAAGAAAAC
                                                                       20195
      TAGAAGGGTA AGAGACCGGT CAGAAAGTAG GCTATTTGAA GTTAACACTT CAGAGGCAGA
                                                                       20255
      GTAGTTCTGA ATGGTAACAA GAAATTGAGT GTGCCTTTGA GAGTAGGTTA AAAAACAATA
                                                                       20315
      GGCAACTTTA TTGTAGCTAC TTCTGGAACA GAAGATTGTC ATTAATAGTT TTAGAAAACT
                                                                       20375
      AAAATATATA GCATACTTAT TTGTCAATTA ACAAAGAAAC TATGTATTTT TAAATGAGAT
                                                                       20435
15
      TTAATGTTTA TTGTAG AA AAC CTG GAA TCA GAT TAC TTT GGC AAG CTT GAA
                       Glu Asn Leu Glu Ser Asp Tyr Phe Gly Lys Leu Glu
      TCT AAA TTA TCA GTC ATA AGA AAT TTG AAT GAC CAA GTT CTC TTC ATT
                                                                       20534
      Ser Lys Leu Ser Val Ile Arg Asn Leu Asn Asp Gln Val Leu Phe Ile
                                     15
20
                                                        20
      GAC CAA GGA AAT CGG CCT CTA TTT GAA GAT ATG ACT GAT TCT GAC TGT
                                                                       20582
      Asp Gln Gly Asn Arg Pro Leu Phe Glu Asp Met Thr Asp Ser Asp Cys
                                 30
                                                    35
             GT ATTTTTTTA ATTCGCAAAC ATAGAAATGA CTAGCTACTT CTTCCCATTC 20638
      AGA G
      Arg Asp
25
          40
      TGTTTTACTG CTTACATTGT TCCGTGCTAG TCCCAATCCT CAGATGAAAA GTCACAGGAG
                                                                      20698
      TGACAATAAT TTCACTTACA GGAAACTTTA TAAGGCATCC ACGTTTTTTA GTTGGGGTAA
                                                                      20758
      AAAATTGGAT ACAATAAGAC ATTGCTAGGG GTCATGCCTC TCTGAGCCTG CCTTTGAATC
                                                                       20818
      ACCAATCCCT TTATTGTGAT TGCATTAACT GTTTAAAACC TCTATAGTTG GATGCTTAAT
                                                                       20878
      CCCTGCTTGT TACAGCTGAA AATGCTGATA GTTTACCAGG TGTGGTGGCA TCTATCTGTA
                                                                      20938
30
      ATCCTAGCTA CTTGGGAGGC TCAAGCAGGA GGATTGCTTG AGGCCAGGAC TTTGAGGCTG 20998
      TAGTACACTG TGATCGTACC TGTGAATAGC CACTGCACTC CAGCCTGGGT GATATACAGA
      TCTACTGTGC CTTCCAAAAC ATGAATTCCA AATATCAAAG TTAGGCTGAG TTGAAGCAGT
                                                                      21178
      GAATGTGCAT TCTTTAAAAA TACTGAATAC TTACCTTAAC ATATATTTTA AATATTTTAT
                                                                      21238
      TTAGCATTTA AAAGTTAAAA ACAATCTTTT AGAATTCATA TCTTTAAAAT ACTCAAAAAA
35
                                                                      21298
      GTTGCAGCGT GTGTGTTGTA ATACACATTA AACTGTGGGG TTGTTTGTTT GTTTGAGATG
                                                                      21358
      21418
      ACTACAACCT CCACCTCCCA CGTTCAAGCG ATTCTCATGC CTCAGTCTCC CGAGTAGGTG
                                                                     21478
      GGATTACAGG CATGCACCAC TTACACCCGG CTAATTTTTG TATTTTTAGT AGAGCTGGGG 21538
      TTTCACCATG TTGGCCAGGC TGGTCTCAAA CCCCTAACCT CAAGTGATCT GCCTGCCTCA 21598
40
      GCCTCCCAAA CAAACAAACA ACCCCACAGT TTAATATGTG TTACAACACA CATGCTGCAA 21658
      CTTTTATGAG TATTTAATG ATATAGATTA TAAAAGGTTG TTTTTAACTT TTAAATGCTG 21718
      GGATTACAGG CATGAGCCAC TGTGCCAGGC CTGAACTGTG TTTTTAAAAA TGTCTGACCA 21778
      GCTGTACATA GTCTCCTGCA GACTGGCCAA GTCTCAAAGT GGGAACAGGT GTATTAAGGA 21838
      CTATCCTTTG GTTAAATTTC CGCAAATGTT CCTGTGCAAG AATTCTTCTA ACTAGAGTTC 21898
      TCATTTATTA TATTTATTTC AG AT AAT GCA CCC CGG ACC ATA TTT ATT ATA
45
                              Asp Asn Ala Pro Arg Thr Ile Phe Ile Ile
                              40
                                                 45
      AGT ATG TAT AAA GAT AGC CAG CCT AGA GGT ATG GCT GTA ACT ATC TCT
                                                                      21997
      Ser Met Tyr Lys Asp Ser Gln Pro Arg Gly Met Ala Val Thr Ile Ser
                         55
                                           60
      GTG AAG TGT GAG AAA ATT TCA ACT CTC TCC TGT GAG AAC AAA ATT ATT
50
                                                                      22045
      Val Lys Cys Glu Lys Ile Ser Thr Leu Ser Cys Glu Asn Lys Ile Ile
                     70
                                        75
```

			AGCCTTACTT	TGTTTTCAAT	CATGTTAATA	TAATCAATAT	22103
	Ser Phe Ly:				•		
5	AATTAGAAAT	ATAACATTAT	TTCTAATGTT	AATATAAGTA	ATGTAATTAG	AAAACTCAAA	22163
•	TATCCTCAGA	CCAACCTTTT	GTCTAGAACA	GAAATAACAA	GAAGCAGAGA	ACCATTAAAG	22223
	TCAATACTTA	CTAAAAATTA	TCAAACTCTT	TACCTATTGT	CATAATGATG	GTTTTTCTGA	22283
	GCCTGTCACA	GGCGAAGAGG	AGATACAACA	CTTGTTTTAT	GACCTGCATC	TCCTGAACAA	22323
	TCAGTCTTTA	TACAAATAAT	AATGTAGAAT	ACATATGTGA	GTTATACATT	TAAGAATAAC	22403
	ATGTGACTTT	CCAGAATGAG	TTCTGCTATG	AAGAATGAAG	CTAATTATCC	TTCTATATTT	22463
10						CTGATCCTTA	22523
	GCCTAAGTCT	TAGACACAAG	CTTCAGCTTC	CAGTTGATGT	ATGTTATTTT	TAATGTTAAT	22583
	CTAATTGAAT	AAAAGTTATG	AGATCAGCTG	TAAAAGTAAT	GCTATAATTA	TCTTCAAGCC	22643
						TATTCTCTAT	22703
	TATTTTTCTC	TATTTCCTCC	ATTATTGTTA	GATAAACCAC	AATTAACTAT	AGCTACAGAC	22763
	TGAGCCAGTA	AGAGTAGCCA.	GGGATGCTTA	CAAATTGGCA	ATGCTTCAGA	GGAGAATTCC	22823
15	ATGTCATGAA	GACTCTTTT	GAGTGGAGAT	TTGCCAATAA	ATATCCGCTT	TCATGCCCAC	22893
						GGGGCAACTT	22943
						ATATGCAACA	23003
	GATACAGCCC	CCAGACAAAT	CCCTCAGCTA	TCTCCCTCCA	ACCAGAGTGC	CACCCCTTCA	23053
	GGTGACAATT	TGGAGTCCCC	ATTCTAGACC	TGACAGGCAG	CTTAGTTATC	AAAATAGCAT	23123
						CTGAACAACA	23183
20						AAACTCAGCT	23243
						TTGTTGAGGG	23303
						GAACTGGGCA	23353
	AATCCATATT	TGGGGGAGCC	TGAAGTTTAT	TCAATTTTGA	TGGCCCTTTT	AAATAAAAAG	23423
	AATGTGGCTG	GGCGTGGTGG	CTCACACCTG	TAATCCCAGC	ACTTTGGGAG	GCCGAGGGG	23433
25	GCGGATCACC	TGAAGTCAGG	AGTTCAAGAC	CAGCCTGACC	AACATGGAGA	AACCCCATCT	23543
	CTACTAAAAA	TACAAAATTA	GCTGGGCGTG	GTGGCATATG	CCTGTAATCC	CAGCTACTCG	23603
		GCAGGAGAAT					23653
	CGTGCCATTG	CACTCCAGCC	TGGGCAACAA	GAGCAAAACT	CGGTCTCAAA	AAAAAAAAA	23723
						TAAGTAGGC	23733
	A DC A DA CC A A	TGGAAGAGAT	CTGTGTAAAT	GAGGGAATCT	GACATTTAAG	CTTCATCAGC	23843
30	CMCACCCMC	CACTACCACCA	CGAAGGAACT	CAATAAATAT	TAGTTGGAGG	GGGGGAGAGA	23903
	UT GAGGGGTG	ACCACMCCMM	AGTTTTAGCC	CTTGTCTTTA	ATCCCTTTC	CTGCCACTAA	23963
		TGGCTCATGC				AT'ACAACAGG	24023
						TCTCTACTAA	24083
						CTACTCGTGA	24143
		GAAGAATCGC					24203 24263
35						GAAAAAGATA	24323
		CCCTTATGTG					24323
						AGGTTACTGT	24443
		TTCATATATA					24503
		TCAGACTTTC					24563
40						GAAATCCCTT	24623
70		GCACTCAAGT					24683
	TCATTGGGAT	TGCTTTCACA	TCCATTTGCT	ATGTACCTTC	CCTATGATGG	CTGGGAGTGG	24743
	TCAACATCAA	AACTAGGAAA	GCTACTGCCC	AAGGATGTCC	TTACCTCTAT	TCTGAAATGT	24803
	GCAATAAGTG	TGATTAAAGA	GATTGCCTGT	TCTACCTATC	CACACTCTCG	CTTTCAACTG	24863
45	TAACTTTCTT	TTTTTCTTTT	TTTCTTTTT	TCTTTTTTT	TGAAACGGAG	TCTCGCTCTG	24923
	TCGCCCAGGC	TAGAGTGCAG	TGGCACGATC	TCAGCTCACT	GCAAGCTCTG	CCTCCCGGGT	24983
	TCACGCCATT	CTCCTGCCTC	ACCCTCCCAA	GCAGCTGGGA	CTACAGGCGC	CTGCCACCAT	25043
	GCCCAGCTAA	TTTTTTGTAT	TTTTAGTAGA	GACGGGGTTT	CACCGTGTTA	GCCAGGATGG	25103
	TCTCGATCTC	CTGAACTTGT	GATCCGCCCG	CCTCAGCCTC	CCAAAGTGCT	GGGATTACAG	25163
	GCGTGAGCCA	TCGCACCCGG	CTCAACTGTA	ACTTTCTATA	CTGGTTCATC	TTCCCCTGTA	25223
	ATGTTACTAG	AGCTTTTGAA	GTTTTGGCTA	TGGATTATTT	CTCATTTATA	CATTAGATTT	25283
50	CAGATTAGTT	CCAAATTGAT	GCCCACAGCT	TAGGGTCTCT	TCCTAAATTG	TATATTGTAG	25343
	ACAGCTGCAG	AAGTGGGTGC	CAATAGGGGA	ACTAGTTTAT	ACTTTCATCA	ACTTAGGACC	25403
						<del>-</del>	

```
CACACTTGTT GATAAAGAAC AAAGGTCAAG AGTTATGACT ACTGATTCCA CAACTGATTG 25463
       AGAAGTTGGA GATAACCCCG TGACCTCTGC CATCCAGAGT CTTTCAGGCA TCTTTGAAGG 25523
       ATGAAGAAAT GCTATTTTAA TTTTGGAGGT TTCTCTATCA GTGCTTAGGA TCATGGGAAT 25583 CTGTGCTGCC ATGAGGCCAA AATTAAGTCC AAAACATCTA CTGGTTCCAG GATTAACATG 25643
5
       GAAGAACCTT AGGTGGTGCC CACATGTTCT GATCCATCCT GCAAAATAGA CATGCTGCAC
                                                                            25703
       TAACAGGAAA AGTGCAGGCA GCACTACCAG TTGGATAACC TGCAAGATTA TAGTTTCAAG
       TAATCTAACC ATTTCTCACA AGGCCCTATT CTGTGACTGA AACATACAAG AATCTGCATT
                                                                            25823
       TGGCCTTCTA AGGCAGGGCC CAGCCAAGGA GACCATATTC AGGACAGAAA TTCAAGACTA 25883
       CTATGGAACT GGAGTGCTTG GCAGGGAAGA CAGAGTCAAG GACTGCCAAC TGAGCCAATA 25943
       CAGCAGGCTT ACACAGGAAC CCAGGGCCTA GCCCTACAAC AATTATTGGG TCTATTCACT 26003
10
       GTAAGTTTTA ATTTCAGGCT CCACTGAAAG AGTAAGCTAA GATTCCTGGC ACTTTCTGTC 26063
       TCTCTCACAG TTGGCTCAGA AATGAGAACT GGTCAGGCCA GGCATGGTGG CTTACACCTG 26123
       GAATCCCAGC ACTTTGGGAG GCCGAAGTGG GAGGGTCACT TGAGGCCAGG AGTTCAGGAC
       CAGCTTAGGC AACAAAGTGA GATACCCCCT GACCCCTTCT CTACAAAAAT AAATTTTAAA
       AATTAGCCAA ATGTGGTGGT GTATACTTAC AGTCCCAGCT ACTCAGGAGG CTGAGGCAGG
                                                                            26303
15
       GGGATTGCTT GAGCCCAGGA ATTCAAGGCT GCAGTGAGCT ATGATTTCAC CACTGCACTT
                                                                           26363
       CTGGCTGGGC AACAGAGCGA GACCCTGTCT CAAAGCAAAA AGAAAAAGAA ACTAGAACTA
                                                                           26423
       GCCTAAGTTT GTGGGAGGAG GTCATCATCG TCTTTAGCCG TGAATGGTTA TTATAGAGGA CAGAAATTGA CATTAGCCCA AAAAGCTTGT GGTCTTTGCT GGAACTCTAC TTAATCTTGA
       GCAAATGTGG ACACCACTCA ATGGGAGAGG AGAGAAGTAA GCTGTTTGAT GTATAGGGGA
                                                                            26603
       AAACTAGAGG CCTGGAACTG AATATGCATC CCATGACAGG GAGAATAGGA GATTCGGAGT
20
                                                                            26663
       TAAGAAGGAG AGGAGGTCAG TACTGCTGTT CAGAGATTTT TTTTATGTAA CTCTTGAGAA
                                                                            26723
       GCAAAACTAC TTTTGTTCTG TTTGGTAATA TACTTCAAAA CAAACTTCAT ATATTCAAAT
                                                                            26783
      TGTTCATGTC CTGAAATAAT TAGGTAATGT TTTTTTCTCT ATAG GAA ATG AAT CCT
                                                                            26839
                                                         Glu Met Asn Pro
                                                         85
      CCT GAT AAC ATC AAG GAT ACA AAA AGT GAC ATC ATA TTC TTT CAG AGA
25
                                                                           26887
      Pro Asp Asn Ile Lys Asp Thr Lys Ser Asp Ile Ile Phe Phe Gln Arg
                              95
                                                 100
      AGT GTC CCA GGA CAT GAT AAT AAG ATG CAA TTT GAA TCT TCA TCA TAC
                                                                           26935
      Ser Val Pro Gly His Asp Asn Lys Met Gln Phe Glu Ser Ser Ser Tyr
                           110
                                               115
      GAA GGA TAC TTT CTA GCT TGT GAA AAA GAG AGA GAC CTT TTT AAA CTC
30
                                                                           26983
      Glu Gly Tyr Phe Leu Ala Cys Glu Lys Glu Arg Asp Leu Phe Lys Leu
                      125
                                          130
                                                                135
      ATT TTG AAA AAA GAG GAT GAA TTG GGG GAT AGA TCT ATA ATG TTC ACT
                                                                           27031
      Ile Leu Lys Lys Glu Asp Glu Leu Gly Asp Arg Ser Ile Met Phe Thr
                  140
                                      145
                                                           150
      GTT CAA AAC GAA GAC T AGCTATTAAA ATTTCATGCC GGGCGCAGTG GCTCACGCCT
35
                                                                           27087
      Val Gln Asn Glu Asp
              155
      GTAATCCCAG CCCTTTGGGA GGCTGAGGCG GGCAGATCAC CAGAGGTCAG GTGTTCAAGA
                                                                           27147
      CCAGCCTGAC CAACATGGTG AAACCTCATC TCTACTAAAA ATACAAAAAA TTAGCTGAGT
                                                                           27207
      GTAGTGACCC ATGCCCTCAA TCCCAGCTAC TCAAGAGGCT GAGGCAGGAG AATCACTTGC
                                                                           27267
      ACTCCGGAGG TGGAGGTTGT GGTGAGCCGA GATTGCACCA TTGCGCTCTA GCCTGGGCAA
                                                                           27327
      CAACAGCAAA ACTCCATCTC AAAAAATAAA ATAAATAAAT AAACAAATAA AAAATTCATA
      ATGTGAACTG TCTGAATTTT TATGTTTAGA AAGATTATGA GATTATTAGT CTATAATTGT AATGGTGAAA TAAAATAAAT ACCAGTCTTG AAAAACATCA TTAAGAAATG AATGAACTTT
      27567
      TAATGTTTAA AAAATTCATA GTTTGAAAAC ATTCTACATT GTTAATTGGC ATATTAATTA
45
      TACTTAATAT AATTATTTT AAATCTTTTG GGTTATTAGT CCTAATGACA AAAGATATTG
      ATATTTGAAC TTTCTAATTT TTAAGAATAT CGTTAAACCA TCAATATTTT TATAAGGAGG
                                                                           27747
      CCACTTCACT TGACAAATTT CTGAATTTCC TCCAAAGTCA GTATATTTTT AAAATTCAGT
                                                                           27807
      50
      ATCATCTCAA TAGATGCAGA AAAAGCATTT GACAAAATCC AATTCATAAT AAAAATTCTC 28047
```

	***	0101101-01					
	AAACTTGAAA	GAGAACATCA	TAAAGGCATC	TATGAAAAAC	CTACAGCTAA	TATCATACTT	28107
	AACGATGAAA	AACTGAATTA	TTTTACCCTA	AGATCAAGAA	TAATGCAAGC	ATCTCACCTC	28167
	TTGCAACTTC	TATTCAACAT	TGTACTGGAG	GTTCTAGCCA	GAGCAACCAT	ACAATAAATA	28227
5	AAAATAAAAG	GCACCCAGAT	TAGAAAGGAA	GTCTTTATTT	GCAGACAACA	TGGTTGTTTA	28287
	TGCAGAAAAC	CGTCAGGAAT	ACACACACAT	GTTAGAACTA	ATAAGTTCAG	CAAGGTTGCA	28347
	GGTTGCAATA	TCAATATGCA	AAAATACATT	GAAGGCTGGG	CTCAGTGGAG	ATGGCATGTA	28407
	CCTTTCGTCC	CAGCTACTTG	GGAGGCTGAG	GTAGGAGGAT	CACTTGAGGT	GAGGAGTTTG	28467
	AGGCTATAGT	GCAATGTGAT	CTTGCCTGTG	AATAGCCACT	GCACTCGAGC	CTAGGCAACA	28527
	AAGTGAGACC	CCGTCTCCAA	AAAAAAAAT	GGTATATTGG	TATTTCTGTA	TATGAACAAT	28587
10	GAATGATCTG	AAAACAAGAA	AATTCCATTC	ACGATGGTAT	TAAAAAAATA	AAATACAAAT	28647
	AAATTTAGCA	AAATAATTAT	AAAACTTGTA	CATCGAAAAT	TTCAAAGCAC	TCTGAGGGAA	28707
	ATTAAAGATG	ATCTAAATAA	TTGGAGAGAC	ACTCTATGAT	CACTGATTGG		28767
	CAATATTGTT	AAGATAACAA	TTGTCCCCAA	ATTGATGCAT	GCATTCAATT	TACTO TO TO TO	28827
	CAAAATTCCA	GC A GG GTTTTT	TOTOTOTOTA	CACAACCMCM	ACCCAAAATG	TAGICTICAT	
		GCNOOGITII	IGCAGAAAII	GACAAGCIGI	ACCCAAAATG	TATATGGAAA	28887
	TGAAAAGACC	CAGAAGAGCA	AATAATTTTT	TAAAAACAAA	GTTGGAAAAC	TTTTACTTCC	28947
13	TAATTTTAAA	ACTTACTATA	AACCTAAAGT	TATCAAGACC	ATTTAGT		28994.

- 14. The genomic DNA of claim 1, which is derived from human.
- 15. The genomic DNA of claim 1, which is inserted into an autonomously replicable vector.
- 16. A transformant derived from a mammalian cell, which contains the genomic DNA claim 1.
- 25 17. The transformant of claim 16, which is derived from a cell selected from the group consisting of epithelial, interstitial and hemopoietic cells from mammal.
  - **18.** A process for preparing a polypeptide, which comprises (a) artificially expressing the DNA of claim 1, and (b) collecting a polypeptide capable of inducing the production of interferon-γ by immunocompetent cells from the resultant mixture.
  - 19. The process of claim 18, wherein the artificial expression of the step (a) comprises a step of culturing the transformant of claim 16.
- 35 20. The process of claim 18, wherein the resultant mixture of the step (b) contains a culture of the transformant of claim 16.
  - 21. The process of claim 18, wherein the polypeptide is collected by one or more techniques selected from the group consisting of salting out, dialysis, filtration, concentration, separatory sedimentation, ion-exchange chromatography, gel filtration chromatography, adsorption chromatography, isoelectricpoint chromatography, hydrophobic chromatography, reversed phase chromatography, affinity chromatography, gel electrophoresis and isoelectric focusing.
- **22.** The process of claim 18, wherein the polypeptide is collected by an immunoaffinity chromatography with a monoclonal antibody.

50

20

30

40

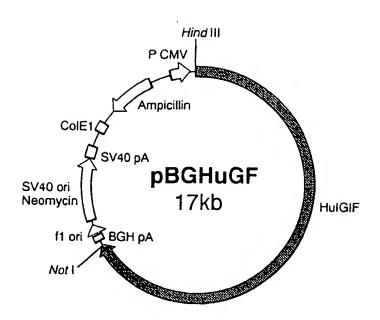


FIG.1

# This Page is Inserted by IFW Indexing and Scanning Operations and is not part of the Official Record

## **BEST AVAILABLE IMAGES**

Defective images within this document are accurate representations of the original documents submitted by the applicant.

Defects in the images include but are not limited to the items checked:

BLACK BORDERS

IMAGE CUT OFF AT TOP, BOTTOM OR SIDES

FADED TEXT OR DRAWING

BLURRED OR ILLEGIBLE TEXT OR DRAWING

SKEWED/SLANTED IMAGES

COLOR OR BLACK AND WHITE PHOTOGRAPHS

GRAY SCALE DOCUMENTS

LINES OR MARKS ON ORIGINAL DOCUMENT

REFERENCE(S) OR EXHIBIT(S) SUBMITTED ARE POOR QUALITY

# IMAGES ARE BEST AVAILABLE COPY.

**☐** OTHER: \_\_\_\_\_

As rescanning these documents will not correct the image problems checked, please do not report these problems to the IFW Image Problem Mailbox.